

Da transparência à opacidade

Autoria: Ana Carolina Faleiros Camargo Moreno
Revisão: Haydée Svab e Fernanda Campagnucci

APRESENTAÇÃO

O presente relatório traça o *histórico* da mudança que o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) realizou em 2022 na política de transparência e divulgação dos microdados educacionais.

Este documento também lista em detalhes as *justificativas* da autarquia do Ministério da Educação (MEC) para a mudança, além de reunir as principais *repercussões e impactos* que ela teve em diversos setores da sociedade civil e em diferentes esferas de governo.



SUMÁRIO

Pág. 4 1. HISTÓRICO

Pág. 5 1.1. Microdados ‘simplificados’

Pág. 8 1.2. Nova política

Pág. 10 1.3. Evolução

Pág. 27 2. JUSTIFICATIVA

Pág. 28 2.1. Risco de reidentificação

Pág. 32 2.2. Episódios em que o risco se concretizou

Pág. 36 2.3. Educação é alvo preferencial de ataques

Pág. 38 2.4. Riscos subsequentes à segurança física dos indivíduos

Pág. 40 2.5. Confiança na privacidade para garantir a qualidade dos dados

Pág. 43 2.6. O Sedap como alternativa à produção de pesquisa

Pág. 46 2.7. Os censos educacionais são “pesquisas estatísticas” e não “registros administrativos”

Pág. 52 2.8. Conclusão

Pág. 54 3. REPERCUSSÃO

Pág. 54 3.1. Sociedade civil

Pág. 54 » Posicionamento

Pág. 56 » Produção de conhecimento

Pág. 58 » Promoção de debate público

Pág. 61 3.2. Controle externo

Pág. 65 3.3. Poder Legislativo

Pág. 65 » Legislação

Pág. 67 » Audiências públicas

Pág. 72 3.4. Poder Judiciário

Pág. 72 3.5. Poder Executivo

Pág. 72 » Inep

Pág. 74 » Autoridade Nacional de Proteção de Dados (ANPD)

Pág. 75 » Controladoria-Geral da União (CGU)

Pág. 79 4. IMPACTO

Pág. 79 4.1. Publicação de artigos e pesquisas acadêmicas

Pág. 85 4.2. Pedidos via Lei de Acesso à Informação

Pág. 88 4.3. Uso dos microdados em geral

Pág. 90 4.4. Atuação da sociedade civil

Pág. 92 4.5. Execução de políticas públicas

Pág. 93 4.6. Limitações do Sedap enquanto alternativa

1. HISTÓRICO

Em 18 de fevereiro de 2022, mais de dez conjuntos de microdados mantidos publicamente pelo Inep foram retirados do site oficial. Entre eles estavam bases de dados como as do Censo Escolar, do Censo da Educação Superior, do Exame Nacional do Ensino Médio (Enem) e do Sistema de Avaliação da Educação Básica, o Saeb, a partir do qual o Inep calcula o Índice de Desenvolvimento da Educação Básica (Ideb), principal indicador de qualidade da educação do Brasil.

A data marcou o fim da política de abertura e transparência de dados educacionais que o instituto expandiu a partir de 2007, tornando-se referência no governo federal para a divulgação de microdados.

Desde então, o atual retrocesso na transparência ativa do Inep teve inúmeros impactos e mobilizou diversos setores, entre universidades, organizações da sociedade civil e representantes dos poderes Executivo, Legislativo e Judiciário, como veremos a seguir.

ENTENDA MAIS

O QUE SÃO MICRODADOS?

Dados são conjuntos de informações produzidos a partir de algum tipo de coleta sobre algum tipo de fenômeno, evento ou registro. Eles podem ser agregados ou não. Os **microdados** são conjuntos de dados no **nível mais desagregado possível**, sem nenhum tipo de consolidação.¹ Por exemplo, uma tabela onde cada carro ocupa uma linha, e cada coluna traz um detalhe sobre o carro, como o modelo, a cor, a rua pela qual ele passou, incluindo a data e a hora. Aqui, cada carro é o que chamamos de **unidade de observação**.

A partir dos microdados (as informações sobre cada carro), é possível elaborar novos conjuntos de **dados agregados**, como uma tabela com o total de carros que passa em cada rua de uma cidade em um determinado dia, incluindo a média de carros por hora de cada rua. Agora, cada rua representa uma unidade de observação.

No caso da Educação, microdados podem ter como unidade de observação os alunos — ou seja, cada linha tem informações sobre um estudante, com as colunas trazendo dados sobre suas características. Os dados podem ser agrupados por turma, escola, município etc., aumentando seu nível de agregação.

1.1. Microdados ‘simplificados’

Quatro dias após tirar os dados do ar, em 22 de fevereiro de 2022, o Inep divulgou um comunicado informando que estava fazendo uma “adequação dos microdados disponíveis em seu portal”. As justificativas para a decisão serão mencionadas mais adiante, mas a principal delas foi “a adequação dos microdados à Lei Geral de Proteção de Dados”.²

Em 31 de março de 2022, o Inep começou a republicar as bases, entretanto em um novo formato, que a autarquia batizou de “microdados simplificados”.

A primeira base a voltar ao ar, nesse novo formato, foi a do Censo Escolar do Inep (também chamado de Censo da Educação Básica). Trata-se de um conjunto de tabelas produzidas a partir da coleta feita em cada instituição de ensino. Anualmente, no mês de maio, todas as escolas públicas e privadas são obrigadas por lei a enviar para o Inep informações específicas de três unidades de observação:

- 1. Da escola**

(como o nome da escola, código da escola, endereço da escola, quantidade de banheiros na escola, entre outros)

- 2. De cada estudante com matrícula na escola**

(como idade, sexo ou cor/raça)

- 3. De cada docente dando aulas na escola**

(sexo, cor/raça, nível de formação etc.)

A partir dessa coleta, o Inep elabora uma base de dados específica de todas as escolas do Brasil, outra dos alunos e uma terceira dos professores. Há ainda uma quarta base de dados, reunindo informações agrupadas para cada **turma** de cada escola. Todo ano, é feita uma nova edição do Censo Escolar.

No entanto, existem várias formas de compilar, tratar e divulgar esses dados. Mantendo o exemplo do Censo Escolar, o Inep historicamente fazia três grandes divulgações:

» **O resumo técnico**

um relatório em PDF contendo textos e gráficos que condensam as principais informações daquela edição do Censo, como total de matrículas por rede de ensino ou por Unidade da Federação;

» **A sinopse estatística**

um arquivo no formato Excel contendo dezenas de abas com tabelas para as principais informações calculadas por município (como a quantidade de matrículas de cada ano ou etapa de ensino);

» **E os microdados**

um arquivo em formato ZIP contendo uma série de outros arquivos em formato CSV reunindo, em cada um, os detalhes anonimizados de cada **escola, estudante, docente e turma**, além de documentos em formato PDF contendo as fichas com os campos que cada escola deve preencher para enviar os dados do Censo, e também informações técnicas para auxiliar no manuseio dos microdados, incluindo filtros a serem aplicados nas tabelas para se chegar aos resultados oficiais agregados e divulgados na sinopse e no resumo estatístico.

Olhando para os níveis de detalhamento acima, chegamos à conclusão de que tanto a sinopse estatística quanto o resumo técnico são produtos de dados agregados, elaborados a partir dos microdados.

No entanto, nos sites de internet do Inep, todos esses dados, independentemente do nível de desagregação, estavam disponíveis na versão **anonimizada**. Ou seja, eram uma versão passível de distribuição pública, depois de passar por um tratamento para a retirada de **dados pessoais**.

ENTENDA MAIS

DADOS PESSOAIS E DADOS SENSÍVEIS

Uma base de dados geralmente é composta por uma grande variedade de informações. Por exemplo, uma base de dados de carros pode conter dados genéricos, como a marca da fabricante, o modelo de veículo, o ano de fabricação e a cor. Mas ela também pode trazer informações mais específicas, como o nome completo do proprietário, além de seu CPF, RG, endereço residencial, endereço de e-mail ou número de telefone pessoal.

Essas informações que podem identificar ou expor um indivíduo são classificadas como **dados pessoais**. Dentro do grupo de dados pessoais, há ainda dois subgrupos que merecem atenção.

O primeiro deles são os chamados **dados sensíveis**, e representam informações sobre uma pessoa que podem expô-la a atos discriminatórios ou que controlem ou restrinjam sua liberdade.³ Alguns exemplos de dados sensíveis são a cor/raça, a orientação sexual ou suas convicções religiosas. Esses dados podem ser coletados, desde que com consentimento do indivíduo e para um uso específico.

Por fim, quando o **indivíduo tem menos de 18 anos**, seus dados pessoais também são sensíveis e só podem ser coletados com consentimento de um dos pais ou responsável legal, ou para usos específicos e sem previsão de armazenamento.⁴

É importante ressaltar que, de acordo com a Lei Geral de Proteção de Dados (LGPD), em seu artigo 12, os dados anonimizados não são considerados dados pessoais, salvo quando o processo de anonimização puder ser facilmente revertido:

Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido (BRASIL, Lei nº 13.709, de 14 de agosto de 2018, Cap. II, Art.12, grifo nosso).

1.2. Nova política

A partir de *março de 2022*, a mudança na política de transparência ativa do Inep implicou a retirada permanente do ar das bases de *alunos, docentes e turmas*. Restaram, então, apenas os microdados no nível da *escola*, que receberam o acréscimo de algumas colunas contendo informações agregadas, como o total de matrículas e docentes da instituição.

A nova política, levada a cabo sem aviso prévio ou consulta pública, provocou inúmeras críticas de diferentes atores. Os dados são frequentemente usados por centenas de pesquisadores para a produção acadêmica de evidências científicas para a realização de diagnósticos e análises, permitindo identificar padrões, desigualdades regionais e fatores que influenciam o aprendizado. Isso contribuiu para a formulação de políticas públicas mais eficazes, lançando luz sobre os desafios do Brasil para ampliar o acesso, a permanência e a qualidade do Ensino Básico e Superior do país.

Entidades da sociedade civil também produzem uma série de insumos alimentados pelos dados, como o Portal QEdU,⁵ onde a população consegue acessar facilmente detalhes como total de matrículas e resultados do Ideb de cada escola ou município. Essa plataforma foi, inclusive, citada no próprio Plano de Dados Abertos do Inep para o biênio 2020-2021, como uma das entidades consultadas em oficina temática com usuários das bases de dados da autarquia.⁶

O Portal QEdU é uma das várias iniciativas privadas de produção de conhecimento educacional. Alguns produtos construídos a partir dos microdados divulgados publicamente pelo Inep eram inclusive financiados pelo próprio governo federal. Um exemplo é a Plataforma de Dados Educacionais,⁷ um projeto do Laboratório de Dados Educacionais da Universidade Federal do Paraná (UFPR) realizado com financiamento do Ministério da Educação.

A partir de fevereiro de 2022, a única alternativa de acesso aos dados antes publicamente disponíveis passou a ser o *Serviço de Acesso a Dados Protegidos (Sedap)*, um espaço físico em Brasília onde pessoas previamente autorizadas podem

fazer buscas e filtros na íntegra das bases de dados (a versão não anonimizada, contendo dados pessoais).⁸

O serviço foi criado originalmente em 2014 e, a partir de julho de 2023, o Inep passou a permitir que entidades se cadastrassem para criar **núcleos Sedap em outras localidades** do país, para acesso remoto e controlado às bases de dados.⁹

Inicialmente, a nova proposta só deu frutos em duas cidades, além de Brasília:

- » O primeiro núcleo Sedap foi inaugurado em junho de 2023, um mês antes da publicação da portaria, no campus da **Universidade Federal de Minas Gerais (UFMG), em Belo Horizonte**.¹⁰
- » Em setembro de 2023, uma Sala Segura do Sedap foi inaugurada na sede do **Inspers, em São Paulo**.¹¹

Posteriormente, novas salas foram abertas em São Paulo e a proposta se expandiu para o estado da Paraíba:

- » Em março de 2024, o **Inspers, em São Paulo**, ganhou uma segunda sala segura autorizada pelo Inep.¹²
- » E em julho de 2024, uma terceira sala foi inaugurada em São Paulo, dessa vez na Faculdade de Economia, Administração e Ciências Contábeis da Universidade de São Paulo (**FEA-USP**).¹³
- » Em setembro de 2024, o projeto chegou pela primeira vez ao Nordeste, com um acordo entre Inep e a **Universidade Federal da Paraíba (UFPB)**.¹⁴

Até o fim de 2024, portanto, além da Sala Segura em Brasília, era possível acessar os microdados em cinco locais de três cidades e estados diferentes (MG, PB e SP). As regiões Sul e Norte ainda carecem de um endereço do tipo.

Outra alternativa prometida pela autarquia, mas que ainda não se concretizou, é uma nova **plataforma de acesso a dados**. Em comunicado publicado em maio de 2024, o Inep informou que a proposta está sendo elaborada em parceria com a Rede Nacional de Ensino e Pesquisa (RNP) para “viabilizar o acesso a dados educacionais produzidos pelo Inep sem descumprir as exigências de proteção à

privacidade e outros direitos fundamentais previstos na Lei Geral de Proteção de Dados Pessoais (LGPD)”.¹⁵

Em 4 de novembro de 2024, durante audiência pública da Subcomissão Permanente para tratar do Sistema Nacional de Educação (SNE),¹⁶ o diretor da Diretoria de Avaliação da Educação Básica (Daeb/Inep), Rubens Campos de Lacerda Júnior, informou que “há a previsão de que até o fim do ano essa plataforma esteja disponível para as análises”. Ainda segundo ele, “um grupo foi criado com a sociedade civil organizada para fazer uma primeira crítica a essa plataforma e fazer as primeiras análises desse modelo”.

Mas, até fevereiro de 2025, ainda não havia acesso público a ela. Em resposta a um pedido feito pela Open Knowledge Brasil (OKBR)¹⁶⁻¹ via Lei de Acesso à Informação (LAI), o Inep explicou que a reunião de kickoff do projeto ocorreu em 12 de abril de 2024 e que, até o início de fevereiro, “a plataforma de acesso a dados educacionais está disponível somente no ambiente de testes e homologação, sendo assim, não está disponível para acesso externo em ambiente de produção”.

Além disso, o Inep confirmou que a sociedade civil não havia sido consultada nas etapas do desenvolvimento da plataforma até fevereiro.

“Não há um grupo da sociedade civil organizada para realizar testes e análises na Plataforma, pois, como explicado anteriormente, a Plataforma de acesso a dados educacionais está disponível, neste momento, somente para testes internos” (Inep, via Lei de Acesso à Informação).

1.3. Evolução

A partir de um levantamento em documentos disponíveis publicamente ou obtidos via Lei de Acesso à Informação, além de entrevistas com atores que estão ou já estiveram no Inep e de pessoas envolvidas na formulação de políticas públicas educacionais ou na produção de pesquisa acadêmica na área, foi possível resgatar como evoluiu historicamente a visão majoritária do instituto em relação à transparência, à privacidade e à função dos dados por ele produzidos.

Esse histórico pode ser dividido em três grandes fases entre 2007 e fevereiro de 2022.

» **Fase 1 (2007 a 2013):** ampliação da transparência.

A primeira publicação de estatísticas educacionais no Brasil ocorreu em 1939, mas o Inep só se tornou o órgão federal oficialmente responsável pela elaboração de levantamentos estatísticos educacionais a partir de 1997.¹⁷

O ano que marca uma virada em termos de aprimoramento do Censo Escolar, além da ampliação da difusão dos dados, é 2007. Foi nesse ano que o Inep criou o Sistema Educacenso e fez uma mudança fundamental na coleta de dados: antes, a unidade de observação dos dados era a escola. A partir de 2007, passou a ser o aluno.

Nas palavras do próprio Inep:

O Censo Escolar 2007 foi totalmente informatizado marcando um novo capítulo na história da evolução do Censo Escolar da Educação Básica. A unidade de informação deixou de ser a escola para ser o aluno. A implantação do EDUCACENSO, sistema de Coleta de informações via Internet, foi bem sucedida em todo o país e contamos, agora, com informações mais precisas e ágeis sobre escolas, alunos e docentes e auxiliares da educação infantil (INEP, Relatório de Gestão 2007, 2008, p.48).¹⁸

Além das melhorias no Censo Escolar, o governo federal também tomou medidas proativas para disseminar o uso dos dados produzidos pela autarquia, com o “incentivo à formação de pesquisadores aptos a estudarem os dados da educação brasileira” e “parcerias com instituições de pesquisa interessadas no desenvolvimento de núcleos voltados à pesquisa educacional com foco em política pública”.

A mais marcante delas foi a criação do **Observatório da Educação**, por meio do Decreto 5.803, de 8 de junho de 2006.^{19,20} A finalidade do observatório era “fomentar a produção acadêmica e a formação de recursos pós-graduados em educação, em nível de mestrado e doutorado, por meio de financiamento específico”. E uma das diretrizes determinadas pelo governo federal era “estimular a utilização de dados estatísticos educacionais produzidos pelo Inep como subsídio ao aprofundamento de estudos sobre a realidade educacional brasileira”.

No Relatório de Gestão 2007, o Inep explicou que o programa teria duração de quatro anos e parceria com a Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), para seleção de “grupos de pesquisadores na área educacional, oferecendo financiamento para pesquisas, especialmente bolsas de mestrado e doutorado, cujos projetos foram escolhidos a partir de seleção pública”.

Segundo o jornal O Globo, em 2007 o governo federal anunciou um investimento de R\$ 1,8 milhão a serem repassados naquele ano para financiar 28 projetos de 19 universidades diferentes. Reajustado pela inflação até dezembro de 2024, o valor real do investimento é de **R\$ 2,7 milhões**.²¹

Em 2008, a pedagoga Gabriela Schneider, do Paraná, ingressou no mestrado na Universidade Federal do Paraná (UFPR). Seu objetivo inicial não era fazer uma pesquisa quantitativa. Mas ela acabou convidada a integrar o grupo de pesquisa de pós-graduação financiado pelo Observatório da Educação e aceitou o convite. Acabou fazendo mestrado e doutorado com bolsa de pesquisa do programa e seguiu carreira como pesquisadora de dados educacionais.

Hoje como professora da UFPR e associada à Associação Nacional de Pesquisa em Financiamento da Educação (Fineduca), ela orienta pesquisas na área e é coordenadora do Laboratório de Dados Educacionais da universidade.

“Muitas universidades começaram a se debruçar [sobre os dados]. Foi uma iniciativa do Inep para incentivar o uso dos microdados. Além da melhor disponibilização, tiveram preocupação de incentivar. Várias universidades fizeram grandes projetos”, lembra Gabriela.

De acordo com a Capes, o último edital do Observatório da Educação foi publicado em 2013.²²

» **Fase 2 (2014 a 2017):** ajustes e aprimoramento.

Em 2014, um novo programa de acesso aos dados do Inep foi criado. Inicialmente, ele foi chamado de **Sala Segura**: um recinto físico designado dentro do prédio onde fica a sede da autarquia, em Brasília, para que pesquisadores pudessem acessar as bases de microdados não anonimizadas.

ENTENDA MAIS

ACESSO A DADOS PROTEGIDOS

Esse tipo de acesso é essencial para alguns tipos específicos de pesquisa, como **pesquisas longitudinais**, que analisam um mesmo indivíduo em diversas edições da base de dados.

Ele também é indispensável para a realização de **cruzamentos entre duas bases de dados** diferentes, mas que contenham o mesmo identificador individual (por exemplo, o CPF).

No entanto, por conterem dados pessoais, as bases não podem ficar disponíveis publicamente, sem qualquer mecanismo de proteção. Os núcleos Sedap funcionam como um desses mecanismos.

Por outro lado, o acesso restrito a uma versão não anonimizada dos dados gera limitações consideráveis, incluindo a presença física do pesquisador em um ambiente específico. Lembrando que o objetivo da transparência, segundo a Lei de Acesso à Informação (LAI), é garantir acesso a dados de interesse público. Já a Sala Segura não pode ser encarada como medida de transparência ativa do Inep, pois ela dá acesso a dados sensíveis, que demandam níveis maiores de proteção. Portanto, a existência da Sala Segura não substitui em medida alguma a disponibilização on-line dos microdados anonimizados.

A permissão de acesso a dados pessoais para pesquisa de interesse público é uma exceção prevista pela legislação brasileira, tanto de transparência, como a LAI, quanto pelas normas de privacidade, como a LGPD.

Em 26 de maio de 2014, a Portaria 209 instituiu o chamado **Ambiente Seguro**, em um “perímetro reservado no primeiro subsolo do prédio do Inep”. A chamada Sala Segura faz parte desse ambiente.

Meses depois, em 19 de setembro, a Portaria 467 criou a **Norma de Acesso às Informações Protegidas do Inep**, definindo o acesso a essa sala.²³ Essa foi a primeira vez que um protocolo oficial definia quem poderia solicitar esse acesso (pessoas físicas ou jurídicas que atendessem aos critérios), que documentos eram exigidos, como o currículo da pessoa pesquisadora e seu projeto de pesquisa, além do termo de consentimento, no caso de pesquisas qualitativas.

Ao longo desse relatório discutiremos as diferenças entre a divulgação pública e o acesso controlado aos dados. Mas, antes, vale discutir o contexto em torno do qual essa medida foi tomada. Isso porque a criação da Sala Segura em 2014 coincidiu com outra iniciativa do tipo: a do **Instituto de Pesquisa Econômica Aplicada (Ipea)**.

Em 20 de maio de 2014 (menos de uma semana antes da Portaria 209 do Inep), o Ipea assina a Portaria 78, que “institui a sala de pesquisa em dados sigilosos no Ipea e os procedimentos necessários para tratamento de informações sigilosas”. O espaço físico foi criado para reunir diferentes bases de dados públicas, restritas ou sigilosas para facilitar o acesso controlado a elas por servidores públicos federais.

Ao contrário do Inep, o Ipea não gera dados. Sua produção é de pesquisas e estudos a partir de bases elaboradas por outros órgãos governamentais, com o objetivo de subsidiar a formulação de políticas públicas. Por isso, a disponibilidade das bases dependia da celebração de acordos com os órgãos produtores de cada uma delas.

A tarefa, porém, não se provou simples. Em um texto para discussão publicado um ano depois, em junho de 2015, pesquisadores do próprio Ipea afirmaram que o órgão “enfrenta recorrentes dificuldades para obter microdados produzidos ou tutelados por órgãos da própria administração pública federal que integra”. Segundo eles, “o arcabouço de leis, normas e regulamentos existente não é impedimento ao compartilhamento de informações sigilosas com outros órgãos,

desde que não seja violado o dever de protegê-las”. A limitação, dizem os pesquisadores, advém da “cultura do sigilo no serviço público”:

É essa cultura que faz com que muitos dos gestores públicos tendam a interpretar a lei de forma restritiva, negando o acesso a informações confidenciais, mesmo quando solicitado por outro órgão da mesma esfera da administração pública. Acabam permitindo, de maneira célere, esse acesso apenas a órgãos que detêm poder de polícia, ou dispõem dessa garantia por lei ou por norma infralegal (ANDRADE, Israel de Oliveira; NASCIMENTO, Paulo A. Meyer M. O sigilo em bases de dados sob a tutela da administração pública: o caso Ipea, 2015, p.13).²⁴

Em junho de 2016, o governo federal facilitou a tarefa com o Decreto 8.789, que dispensava a celebração de convênios ou acordos de cooperação técnica no **compartilhamento de bases de dados dentro da administração pública federal**.²⁵ O decreto foi substituído em 2019 pelo Decreto 10.046, que criou o Comitê Central de Governança de Dados, mas manteve a dispensa de acordos específicos para o compartilhamento de bases de dados entre órgãos.²⁶

Mas, no caso do Inep, a reação à criação da sala sigilosa do Ipea foi a movimentação para fazer o mesmo dentro da autarquia, ou seja, sem compartilhar as bases com o público externo. Nem o decreto de 2016 influenciou no compartilhamento dos dados com o Ipea. Prova disso é a aparição, nos registros de uso do Sedap, de pesquisador do Ipea que precisou ir fisicamente à sala do Inep, em 2022, para acessar dados necessários para uma pesquisa sobre as trajetórias ocupacionais de engenheiros e profissionais de Tecnologia da Informação (TI) no Brasil.²⁷

Em resposta a um questionamento feito por e-mail pela Open Knowledge Brasil (OKBR), o Ipea confirmou que “os microdados educacionais produzidos pelo Inep nunca foram compartilhados com o Ipea. Assim, se algum projeto depende desses dados, o Ipea deve recorrer à sala de sigilo do Inep, seguindo os mesmos procedimentos que os outros usuários”.

Por questões de segurança, o Ipea afirma que não divulga publicamente o catálogo de bases de dados disponíveis na sala, mas diz que “novas bases chegam continuamente ao longo do ano, dependendo de Acordos de Cooperação

Técnica (ACTs), demandas de pesquisa dos ministérios e pedidos do Ipea a órgãos produtores ou controladores de dados”.

Segundo fontes ouvidas sob condição de anonimato, há uma grande variedade de conjuntos de dados que, em alguns casos, passaram por pré-tratamentos para torná-los ainda mais fáceis de consultar e analisar. Um exemplo são os dados da Relação Anual de Informações Sociais (Rais), do Ministério do Trabalho e Emprego (MTE), disponíveis de forma mais qualificada que a versão do próprio ministério.

A sala de sigilo, segundo o Ipea, é aberta apenas “para utilização do público interno do Instituto e, eventualmente, por algum ministério”, embora os usuários precisem abrir um processo interno para terem o acesso autorizado, além de assinar um termo de responsabilidade.

Desde a pandemia, a sala conta com uma nova solução de segurança chamada de “servidor de acesso restrito”, que garante o acesso remoto e protegido a dados. Assim, nos últimos anos, “somente bases com alto potencial de dano em caso de vazamento são mantidas apenas na sala de sigilo (bases com informações sobre condições de saúde, relacionadas a atividades de segurança pública), ou bases cujo produtor ou controlador dos dados exigiu que estes fossem mantidos nesse ambiente”.

No Inep, desde a implantação da Sala Segura, o serviço também passou por aprimoramento. Em 2015, conforme o relatório de gestão da autarquia, foi criado o **Serviço de Atendimento ao Pesquisador (SAP)**, para atender solicitações específicas de “consultores, professores e pesquisadores”.²⁸

Em 2016, o órgão publicou sua **Política de Dados Abertos** e aprovou o Plano Institucional de Dados Abertos,²⁹ além de ter criado um **Comitê Permanente de Dados**.³⁰

Ele também instituiu o **Grupo de Trabalho do Serviço de Atendimento ao Pesquisador (GT/SAP)** para estudar e propor melhorias no serviço.

Segundo o relatório de gestão do Inep para aquele ano, a posição da Diretoria de Estatísticas Educacionais (Deed) era a de ampliar o acesso da sociedade às informações:

A Deed tem trabalhado para atender aos padrões (nacionais e internacionais) de divulgação de estatísticas e dados, em especial considerando o padrão de dados abertos, para atender diretamente um maior número de perfis de usuários desse tipo de informações (INEP, Relatório de Gestão do Exercício 2016, Brasília 2017, p.23).³¹

O GT foi concluído em outubro daquele ano e produziu uma série de recomendações, a partir das quais foi oficialmente criado o ***Serviço de Acesso a Dados Protegidos (Sedap)***, em maio de 2017, com a implementação de ajustes legais e administrativos, além de uma melhor padronização dos protocolos de atendimento do serviço.^{32,33}

Ainda em 2017, o Inep publicou, em fevereiro, uma portaria com “os ***princípios fundamentais*** e boas práticas que orientam a produção e divulgação das estatísticas educacionais oficiais produzidas pelo Inep”. Entre as considerações levadas em conta para divulgar o documento, o Inep afirmou “que as estatísticas oficiais constituem-se um instrumento fundamental para a gestão, o acompanhamento e avaliação de políticas públicas e de planos e programas governamentais; portanto elemento indispensável para o sistema de informação de uma sociedade democrática”.³⁴

O documento contém 16 princípios diferentes, com suas definições, motivação e ações. O último deles se chama “***acessibilidade e transparência***” e lista quatro ações:

1. Garantir o acesso livre e igualitário às estatísticas e outros documentos relativos ao processo de sua produção para todos os interessados, por meio de procedimentos públicos claramente estabelecidos.
2. Divulgar de forma clara e precisa as estatísticas e seus respectivos metadados, a fim de facilitar o acesso, a correta interpretação, a realização de comparações úteis e a interoperabilidade com outros sistemas de informação.
3. Avaliar o alcance e a satisfação dos usuários em relação aos diferentes produtos estatísticos disponíveis, utilizando-se dos resultados dessa avaliação para aprimorá-los.

4. Promover as estatísticas educacionais junto aos órgãos de imprensa e usuários em geral.

De acordo com o documento, “as estatísticas educacionais devem ser apresentadas de forma clara e compreensível, e divulgadas adequadamente, permitindo o acesso igualitário a todos os perfis de usuários”. E as ações listadas acima são motivadas pelo fato de que, “para que as estatísticas educacionais sejam efetivas, devem estar disponíveis e acessíveis a todos os interessados”.

Já o item 4 do documento versa sobre a “confidencialidade estatística”, e estipula que “a privacidade dos indivíduos, a proteção das informações individuais e a utilização exclusiva dos dados para as finalidades especificadas em legislação devem ser asseguradas em todo o processo estatístico”, e ressalta que “a percepção de que os dados divulgados permitam a identificação do indivíduo aumenta o risco de respostas não fidedignas, resultando em viés ou erro não controlado”.

Na portaria, o Inep não explicou como priorizou os 16 princípios que norteiam sua produção e divulgação de estatísticas, mas o fato de a transparência ter sido listada como último item, e a confidencialidade ter ficado em quarto lugar, condiz com as medidas que foram tomadas a partir de então.

- » **Fase 3 (a partir de 2018):** eliminação de riscos a qualquer custo.

O principal instrumento jurídico que o Inep usa como base para a mudança na política de divulgação de microdados é a Lei Federal nº 13.709, também conhecida como **Lei Geral de Proteção de Dados (LGPD)**. Ela foi sancionada em agosto de 2018, mas com um longo período de adaptação: a grande maioria dos artigos que determinam a conduta dos servidores públicos federais só entrou em vigor 24 meses depois da sanção.

Durante esse processo de transição, a autarquia do MEC tomou uma série de medidas para se adequar às novas regras de proteção dos dados pessoais. Em fevereiro de 2019, o Inep alterou uma **regra de codificação** de indivíduos nas

versões públicas das bases de dados dos censos da Educação Básica e da Educação Superior. As bases já não continham dados pessoais dos alunos e docentes, mas cada estudante e cada professor tinha um código único (chamado de CO_PESSOA_FISICA), que se repetia nas diferentes edições. Com essa informação, não era possível saber quem era aquele indivíduo, mas era possível achar o mesmo indivíduo em dois anos diferentes do censo, por exemplo (e acessar seus detalhes públicos, como idade, cor/raça, nome da escola, série em que estava matriculado, no caso de estudantes, ou informações sobre um professor que leciona em mais de uma escola em um ano específico).

Com essa alteração, o Inep substituiu o CO_PESSOA_FISICA pelas colunas ID_ALUNO e ID_DOCENTE, respectivamente. Esse ID passou a mudar a cada edição das bases, o que eliminou a possibilidade de cruzar bases de anos diferentes e encontrar o mesmo indivíduo nelas – apesar de não ser possível acessar dados pessoais dele, seria possível verificar se mudou de escola ou acompanhar seu fluxo ao longo dos anos.

Na nota em que comunicou a alteração, o Inep informou que tomou a medida “com o objetivo de garantir a proteção do dado pessoal”. Disse, ainda, que “os Microdados dessas pesquisas passarão, em breve, por outras adequações para atender a todas as disposições legais vigentes relacionadas à proteção do dado pessoal”.³⁵

Em julho de 2019, o Inep também atualizou as regras e normas que gerem o *Sedap*. Essa redação seguia em vigor até junho de 2025.³⁶

Em novembro de 2019, o instituto realizou um *workshop interno* sobre LGPD com a participação dos diretores, do chefe de Gabinete e dos coordenadores-gerais.³⁷

Em abril de 2020, foi publicada a *Política de Segurança da Informação e Comunicações - PoSIC*,³⁸ atualizada em maio de 2021.³⁹ Esse documento estipula que todos os bancos de dados devem ter *procedimentos próprios de segurança*, nos termos da LGPD.

No mesmo mês, o Inep, por meio da Deed, tomou uma das principais medidas: assinou, com a Universidade Federal de Minas Gerais (UFMG), o Termo de Execução Descentralizada (TED) de número 8750, cujo objeto era o “Controle de Privacidade na Divulgação dos Censos Educacionais”.

O TED 8750 foi batizado de Price (Privacidade nos Censos Educacionais) e, de acordo com dados publicados no Diário Oficial da União,⁴⁰ teria vigência entre 29 de abril de 2020 e 15 de dezembro de 2020, com custo de **R\$ 538.755,40**. Porém, em 25 de novembro de 2020, Inep e UFMG assinaram um aditivo ao contrato inicial, subindo o valor total para **R\$ 620.000,00**, com vigência retroativa de 8 de maio de 2020 a 28 de abril de 2021.⁴¹ O valor total pago nesse projeto foi de **R\$ 609.883,83**.

O trabalho foi executado pelo **Laboratório Inscript** (Laboratory of Information Security, Cryptography, Privacy, and Transparency), do Departamento de Ciência da Computação (DCC-UFMG), e resultou em oito produtos diferentes:⁴²

- » **Produto 1:** Relatório sobre o panorama internacional e o contexto do Inep a respeito dos métodos de tratamento de controle de privacidade na divulgação estatística
- » **Produto 2:** Relatório sobre os riscos à privacidade decorrentes da atual forma de divulgação dos microdados dos Censos Educacionais do Inep
- » **Produto 3:** Relatório técnico sobre métodos de tratamento aplicáveis aos microdados de divulgação dos Censos Educacionais do Inep
- » **Produto 4:** Solução tecnológica e sua documentação
- » **Produto 5:** Relatório técnico final do projeto-piloto
- » **Produto 6:** Relatório técnico de implantação
- » **Produto 7:** Relatório de operação assistida
- » **Produto 8:** Relatório de encerramento do projeto

No único material público que o Inep divulgou sobre esse trabalho, é possível acessar os sete primeiros produtos em um documento de 741 páginas. Uma análise desse conteúdo mostra que a equipe executora do TED (da UFMG) e o Inep realizaram, entre 20 de maio de 2020 e 3 de março de 2021, seis reuniões para discutir o estudo.

O documento em si, segundo as páginas finais do produto 7, foi finalizado em abril de 2021, antes da entrega do produto 8 (o relatório de encerramento do projeto) que, até janeiro de 2025, não havia sido disponibilizado pelo Inep na página que reúne as informações sobre a nova política de divulgação dos microdados.

No próximo capítulo, veremos com mais profundidade os achados do estudo que fundamentaram a nova política do Inep de não divulgação dos microdados, a não ser na nova versão “simplificada”. Mas chama a atenção, nas informações iniciais, que todos os produtos do TED trazem a premissa de que o instituto considerava agosto de 2020, o mês de entrada em vigor dos artigos da LGPD, como um divisor de águas. E esse prazo seria o marco a partir do qual a forma de divulgação de microdados existente até então se tornaria incongruente com a legislação.

Em diversos pontos do documento, o estudo ressalta que “o Inep dispõe de um prazo exíguo para implementação de uma solução que atenda às demandas da LGPD” e que:

Dada esta restrição de tempo, o Inep deve considerar adotar uma solução emergencial de curto prazo que garanta a privacidade dos titulares – mesmo que seja possivelmente sub-ótima em relação à utilidade provida aos analistas (ou vice-versa) –, enquanto elabora-se uma alternativa de longo prazo que forneça um melhor custo-benefício na relação de privacidade e utilidade (INEP; UFMG. TED 8750 – PRICE: Privacidade nos Censos Educacionais, 2020, p. 262-263).

Nesse caso, a primeira solução emergencial sugerida foi “eliminar imediatamente a divulgação de microdados em qualquer forma, a fim de garantir a privacidade em nível individual exigida pela LGPD”.

Como sabemos, essa medida só foi tomada quase dois anos depois da entrada em vigor da LGPD, em fevereiro de 2022. Antes de chegar a esse ponto, o Inep tomou novas medidas que ajudaram a construir o caminho para a decisão de retirar os microdados do ar.

No mês de novembro de 2020, foi criada uma **força-tarefa** para realizar, num prazo de três meses, um diagnóstico dos impactos da LGPD no âmbito da

autarquia.⁴³ O trabalho do grupo, que contou com representantes de todas as diretorias do Inep, foi prorrogado mais duas vezes, até 31 de julho de 2021.^{44,45}

Em abril de 2021, uma manifestação do Inep via Serviço de Informação ao Cidadão (SIC) indicou que a autarquia ainda previa manter o calendário de divulgação de microdados. Questionada por um cidadão, ela informou que a “previsão é que os microdados públicos do [Exame Nacional do Ensino Médio] ENEM 2020 estejam disponíveis para *download* no portal do INEP **até o final de julho de 2021**”.

Porém, em seguida **duas notas técnicas** tiveram origem nas principais diretorias do órgão relacionadas à produção de dados: a **Deed**, responsável pela coleta e elaboração dos censos da Educação Básica e da Educação Superior, e a **Diretoria de Avaliação da Educação Básica (Daeb)**, que realiza o Enem anualmente, e produz uma base de dados com informações socioeconômicas e de desempenho dos indivíduos que se inscreveram em cada edição.

A **nota técnica da Deed** foi a primeira a ser publicada, em maio de 2021.⁴⁶ Nela, é apresentado um resumo dos achados e recomendações produzidos pelo TED 8750, além do apontamento de “fragilidades apontadas nos microdados públicos dos Censos Educacionais, e provavelmente nos microdados públicos de outras pesquisas, avaliações e exames, divulgados no site do Inep”.

Veremos, no próximo capítulo, os argumentos listados na nota para evidenciar tais fragilidades. A nota conclui que “a forma adotada atualmente pelo Inep para o tratamento da privacidade nos microdados públicos é ineficaz” e sugere uma série de medidas, como uma análise da **Procuradoria Federal especializada junto ao Inep (Projur)** sobre a adequação legal da disseminação de dados, e solicitar “apoio técnico” da Agência Nacional de Proteção de Dados (ANPD) para o “desenvolvimento de ações intensivas e de curto prazo para o alinhamento das atividades do Inep aos requisitos da LGPD, **incluindo a retirada do ar dos microdados de pesquisas**”.

Além disso, a nota sugere alternativas aos microdados, como “ampliar a capacidade de atendimento do Sedap”, “soluções de disseminação de dados de maneira

dinâmica” (ou seja, a produção de painéis ou *dashboards*/painéis com informações agregadas) e “desenvolvimento de solução remota (não presencial) para o Sedap”. Todas as medidas acima foram adotadas pelo Inep.

Também é proposto “realizar seminário com os diferentes perfis de usuário” para tratar sobre o tema da divulgação dos microdados, “de forma a atender os requisitos legais e relação otimizada entre privacidade e transparência na comunicação científica”. Segundo levantamento feito com entidades da sociedade civil, pesquisadores universitários e ministérios do governo federal, não houve debate prévio a respeito do assunto antes da mudança da política.

Em 28 de julho de 2021, a *Daeb* também se pronunciou em *nota técnica*, por meio de sua *Coordenação-Geral de Instrumentos e Medidas (Cgim)*.⁴⁷ O documento confirma que “os microdados da edição de 2020 do ENEM encontram-se prontos para divulgação, seguindo o mesmo modelo utilizado em edições anteriores”.

No entanto, a nota diz que o modelo “estaria fora de escopo da recente legislação de proteção de dados pessoais”. A Cgim recomenda “a realização de um estudo, similar ao realizado pela Deed em parceria com a UFMG, para avaliar a adequação quanto aos requisitos e vedações trazidos pela LGPD dos produtos informacionais (microdados e sinopses estatísticas) elaborados pela Daeb”.

Enquanto tal estudo não fosse realizado, a coordenação propôs *remover algumas colunas da base de dados do Enem 2020* antes da publicação do arquivo no site, para reduzir riscos de reidentificação de indivíduos. Ela informa, então, que “aguarda orientações” da diretoria, da Projur e da presidência do Inep.

Dois dias após a assinatura da nota técnica da Cgim/Daeb, em 30 de julho de 2021, o Inep dá mais um passo, dessa vez com uma portaria que cria a figura do *Encarregado pelo Tratamento de Dados Pessoais* no âmbito da autarquia.⁴⁸

O cargo foi atribuído ao chefe da *Assessoria de Governança e Gestão Estratégica (AGGE)*, uma área do Inep criada em maio de 2021 e vigente até o momento. Para

auxiliá-lo no trabalho operacional de interlocução com o público e com a ANPD, além da orientação de servidores e tomada de providências, a mesma portaria criou o ***Escritório de Apoio à Proteção de Dados (EAPD)***, inicialmente com vigência de 12 meses, mas podendo ser prorrogado.

Em 2 de agosto do mesmo ano, entra no ar uma página do site do Inep dedicada ao ***tratamento de dados pessoais***, que atribui à chefia da AGGE a função de encarregado pelo tratamento dos dados.⁴⁹

Já em relação à ***Projur***, que foi citada nas duas notas técnicas da Deed e da Daeb, a manifestação só chegou no ano seguinte. O parecer 18/2022⁵⁰ trouxe uma análise jurídica em relação às medidas administrativas de divulgação de microdados educacionais e do TED 8750, além de uma interpretação da LGPD, do risco de identificação dos titulares dos dados, das orientações da ANPD e de precedentes do Supremo Tribunal Federal (STF), finalizando com “recomendações de caráter técnico-jurídico”.

O parecer concluiu que “a divulgação dos dados dos censos educacionais ou exames de competência legal do Inep precisa passar pelo filtro de critérios objetivos que ***reduzam, quando não eliminem, o risco potencial de identificação das pessoas a quem os dados estatísticos se referem nos microdados***”.

Além disso, o documento recomenda que, “se a divulgação dos censos ou outras bases de dados mantidos pelo Inep puder resultar em acesso, por terceiros, a microdados pessoais não anonimizados ou que permitam a reidentificação de seus titulares, a ***divulgação não poderá ser realizada, de acordo com a LGPD***”.

A data de assinatura é de 17 de fevereiro de 2022. No dia seguinte, os microdados educacionais, no formato em que eram divulgados desde 2008, saíram do ar.

Referências

1. Disponível em: <<https://www.cetic.br/microdados/>>.
2. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/nota-de-esclarecimento-divulgacao-dos-microdados>>.
3. DONEDA, Danilo. Privacidade e transparência no acesso à informação pública. In: MEZZARROBA, Orides; GALINDO, Fernando. Democracia eletrônica. Zaragoza: Prensas de la Universidad de Zaragoza, 2010, p. 179-216. https://scholar.google.it/citations?view_op=view_citation&hl=pt-BR&user=gSv2oD4AAAAJ&cstar_t=20&pagesize=80&citation_for_view=gSv2oD4AAAAJ:Tyk-4Ss8FVUC
4. Disponível em: <<https://www.gov.br/mds/pt-br/aceso-a-informacao/lgpd/classificacao-dos-dados>>.
5. Disponível em: <<https://qedu.org.br/>>.
6. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/gestao_e_governanca/politica_e_plano_de_dados_abertos.pdf> (p. 21).
7. Disponível em: <<https://dadoseducacionais.c3sl.ufpr.br/plataforma/sobre>>.
8. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/gestao-do-conhecimento-e-estudos-educacionais/cibec/servico-de-aceso-a-dados-protegidos-sedap>>.
9. Disponível em: <<https://www.in.gov.br/web/dou/-/portaria-n-312-de-3-de-julho-de-2023-494164591>>.
10. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/cgdi/inep-inaugura-nucleo-do-sedap-na-ufmg>>.
11. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/cgdi/nucleo-do-sedap-e-inaugurado-em-sao-paulo>>.
12. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/cgdi/inep-amplia-servico-de-aceso-a-dados-protegidos-no-insper>>.
13. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/cgdi/inep-inaugura-novo-nucleo-do-sedap-em-sp>>.
14. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/cgdi/inep-e-ufpb-oficializam-acordo-para-nucleo-do-sedap>>.
15. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/inep-apresenta-nova-plataforma-de-dados>>.
16. Disponível em: <<https://escriba.camara.leg.br/escriba-servicosweb/html/74637>> (resposta registrada às 12h11)
16-1. PEDIDO DE ACESSO À INFORMAÇÃO, protocolo nº 23546.117501/2024-48, solicitado pela Open Knowledge Brasil, em 22/12/2024.
17. Disponível em: <<https://ces.ibge.gov.br/apresentacao/portarias/200-comite-de-estatisticas-sociais/base-de-dados/1185-censo-escolar-educacao-basica.html>>.
18. Disponível em: <https://download.inep.gov.br/download/institucional/relatorios/Relatorio_Gestao_2007.pdf> (p. 48).
19. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/decreto/d5803.htm>.
20. Disponível em: <<https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/educacao-basica/programas-encerrados/observatorio-da-educacao/observatorio-da-educacao>>.
21. Disponível em: <<https://oglobo.globo.com/brasil/educacao/pesquisas-universitarias-recebem-r-1-8-milhao-4238187>>.
22. Disponível em: <<https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/educacao-basica/programas-encerrados/observatorio-da-educacao/editais-observatorio-da-educacao>>.
23. Disponível em: <<https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?jornal=1&pagina=25&data=22/09/2014>>.

24. Disponível em: <<https://repositorio.ipea.gov.br/handle/11058/4418>> (p. 13).
25. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8789.htm>.
26. Disponível em: <https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D10046.htm#art34>.
27. Disponível em: <https://download.inep.gov.br/cibec/documentos/2022/publicacao_processos_sedap_abertos_2022.pdf>.
28. Disponível em: <https://download.inep.gov.br/download/inep/relatorio_gestao_2015_versao_30.03.16.pdf>.
29. Disponível em: <https://download.inep.gov.br/institucional/legislacao/2016/portaria_n370.pdf>.
30. Disponível em: <https://download.inep.gov.br/download/inep/relatorio_gestao_2017_tcu_anexos_29Mar2018.pdf> (p. 173).
31. Disponível em: <https://download.inep.gov.br/download/inep/relatorio_gestao_2016.pdf> (p. 23).
32. Disponível em: <https://download.inep.gov.br/dados/sedap/portaria/2017/portaria_n465_de_31052017_criacao_sedap.pdf>.
33. Disponível em: <https://download.inep.gov.br/download/inep/relatorio_gestao_2017_tcu_anexos_29Mar2018.pdf> (p. 143).
34. Disponível em: <https://download.inep.gov.br/educacao_basica/censo_escolar/legislacao/2017/portaria_inep_91_02022017_principios_fundamentais_estatisticas_educacionais.pdf>.
35. Disponível em: <https://download.inep.gov.br/microdados/Nota_informativa_versao_01.pdf>.
36. Disponível em: <https://download.inep.gov.br/dados/sedap/portaria/2019/portaria_637_17072019.pdf>.
37. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/gestao_e_governanca/relatorio_de_gestao_2019.pdf> (p. 24).
38. Disponível em: <<https://pesquisa.in.gov.br/imprensa/servlet/INPDFViewer?journal=515&pagina=58&data=20/04/2020&captchafield=firstAccess>>.
39. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/gestao_e_governanca/politica_de_seguranca_da_informacao_e_comunicacoes_posic.pdf>.
40. Disponível em: <<https://pesquisa.in.gov.br/imprensa/servlet/INPDFViewer?journal=530&pagina=43&data=05/05/2020&captchafield=firstAccess>>.
41. Disponível em: <<https://pesquisa.in.gov.br/imprensa/servlet/INPDFViewer?journal=530&pagina=70&data=25/11/2020&captchafield=firstAccess>>.
42. Disponível em: <https://download.inep.gov.br/microdados/TED_8750-UFMG.pdf>.
43. Disponível em: <https://download.inep.gov.br/aceso_a_informacao/tratamento_de_dados_pessoais/portaria_592-2020.pdf>.
44. Disponível em: <https://download.inep.gov.br/aceso_a_informacao/tratamento_de_dados_pessoais/portaria_70-2021.pdf>.
45. Disponível em: <https://download.inep.gov.br/aceso_a_informacao/tratamento_de_dados_pessoais/portaria_198-2021.pdf>.
46. Disponível em: <https://download.inep.gov.br/microdados/nota_tecnica_5-2021_deed.pdf>.
47. Disponível em: <https://download.inep.gov.br/microdados/nota_tecnica_14-2021_daeb.pdf>.
48. Disponível em: <https://download.inep.gov.br/aceso_a_informacao/tratamento_de_dados_pessoais/portaria_281-2021.pdf>.
49. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/tratamento-de-dados-pessoais>>.
50. Disponível em: <https://download.inep.gov.br/microdados/parecer_00018-2022_PFInep.pdf>.

2. JUSTIFICATIVA

Como visto no capítulo anterior, a opção do Inep de retirar os microdados educacionais do ar não foi tomada a partir de um impulso, no improviso ou em atendimento a uma diretriz governamental. Ao longo de vários anos, diversas ações internas da autarquia produziram um rastro de documentação, acumulando argumentos jurídicos e técnicos para embasar a decisão.

Alguns desses argumentos datam de anos antes de agosto de 2018, quando ocorreu a aprovação da Lei Geral de Proteção de Dados (LGPD). Mas eles também se enquadram na listagem de motivos pelos quais o Inep concluiu que a política de divulgação de microdados para a sociedade era incompatível com a nova legislação e, portanto, “ineficaz para o contexto atual”.¹

Vale lembrar que a LGPD estipula claramente que dados anonimizados não representam dados pessoais e, portanto, não estão submetidos ao texto da lei. No entanto, o Inep ressalta que, usando dados externos às bases públicas de microdados educacionais, existe risco de reidentificação de um indivíduo, o que faria com que ele deixasse de ser anônimo nas tabelas divulgadas pelo instituto.

Em síntese, os argumentos que a autarquia ofereceu para a mudança da política são os seguintes:

1. Os microdados, da forma como eram divulgados, oferecem **risco de reidentificação** de indivíduos (estudantes e professores).
2. A **concretização desse risco** já ocorreu em alguns episódios pontuais.
3. O setor de educação é **alvo preferencial de ataques**.
4. Os ataques de reidentificação podem culminar em **riscos à segurança física** dos indivíduos.
5. As famílias só aceitariam fornecer informações para o censo se **confiarem que o Inep protegerá seus dados pessoais**, o que afeta a qualidade final das bases.
6. Retirar os microdados no site **não inviabiliza a produção de pesquisas** porque o Inep oferece a pesquisadores o Serviço de Acesso a Dados Protegidos (Sedap).
7. Os censos educacionais são **“pesquisas estatísticas” e não “registros**

administrativos". Por isso, devem seguir as diretrizes da produção científica, e não submetidos às regras de transparência ativa reservadas aos registros

Neste capítulo, destrinchamos cada um deles, oferecendo, no fim, uma síntese dos contra-argumentos, de acordo com especialistas nas áreas de Direito, Proteção de Dados, Pesquisa, Políticas Públicas e Segurança de Informação.

2.1. Risco de reidentificação

As evidências principais sobre riscos de identificação usadas pelo Inep saíram do Termo de Execução Descentralizada (TED) 8750 assinado entre o Inep e a Universidade Federal de Minas Gerais (UFMG), por meio do Laboratório Inscript (Laboratory of Information Security, Cryptography, Privacy, and Transparency), do Departamento de Ciência da Computação (DCC-UFMG).²

O estudo se restringiu às bases de dados do *Censo Escolar* e do *Censo da Educação Superior*, que são produzidas pela Diretoria de Estudos Educacionais (Deed), e faz uma análise das atuais técnicas aplicadas pelo Inep para proteger os dados pessoais, que são a *desidentificação* e a *pseudonimização*.

ENTENDA MAIS

CONCEITOS DE PRIVACIDADE DE DADOS

Desidentificação: é a técnica pela qual identificadores individuais óbvios são removidos dos registros. Entre esses identificadores estão nome, CPF e RG.

Pseudonimização: nessa técnica, os identificadores individuais óbvios são substituídos por um código único de identificação artificialmente criado.

Identificadores individuais: são as informações únicas de cada indivíduo, ou seja, aquele dado só se remete a uma única pessoa, como o CPF.

Quase-identificadores: são informações que, sozinhas, não permitem identificar um indivíduo (por exemplo, muitas pessoas têm o mesmo CEP, a mesma data de nascimento e o mesmo gênero). Mas, agrupadas com outros quase-identificadores, aumentam a probabilidade de um indivíduo ser reidentificado na base de dados. No exemplo anterior, a chance de existir apenas uma mulher nascida naquela data e morando naquele CEP é muito mais alta.

Para testar a eficácia das técnicas de proteção usadas pelo Inep, o Laboratório Inscript desenhou simulações de diversos “ataques”, como são chamadas as tentativas de quebrar a privacidade individual em bases de dados. São eles:

- » **Ataques de reidentificação:** o objetivo é identificar indivíduos específicos dentro da base de dados anonimizada. Esses ataques podem ser de reidentificação individual ou coletiva.
- » **Ataques de inferência de atributo:** aqui, o objetivo é deduzir o valor de um atributo sensível de um indivíduo, mesmo sem necessariamente reidentificá-lo.
- » **Ataques longitudinais:** são os que usam dados de diferentes edições do Censo para aumentar as chances de reidentificação ou inferência de atributos.

Os ataques partiram de dois pressupostos obrigatórios:

1. O adversário possui acesso irrestrito à **base de dados anonimizada**, ou seja, aquela em que identificadores diretos, como nome e CPF, foram removidos.
2. O adversário **já possui conhecimento prévio** de informações chamadas de auxiliares a respeito dos indivíduos. Não só se conhece o indivíduo em si, mas é preciso já saber características extras sobre ele, como data de nascimento, cidade natal, instituição de ensino e curso.

Ou seja, nenhuma inferência foi estudada sem que dados quase-identificadores obtidos de **fontes externas** fossem usados no ataque.

As informações dos censos educacionais usadas pela UFMG como quase-identificadores foram, entre outras:

- » data de nascimento;
- » município de nascimento;
- » código individual e dependência administrativa da escola;
- » código do curso em que a pessoa está matriculada
- » cor/raça;
- » nacionalidade;
- » país de origem da pessoa.

A partir daí, foram realizados vários ataques diferentes, usando diversas combinações dessas características previamente obtidas.

Os resultados foram quantificados segundo a probabilidade de o adversário reidentificar corretamente um indivíduo ou grupo de indivíduos previamente conhecidos.

Os principais resultados destacados pelo Inep são:

- » **Ataques ao Censo Escolar da Educação Básica 2019:**
 - » O uso de três quase-identificadores (*mês, ano de nascimento e código da escola* em que estuda) permite a identificação com probabilidade de acerto de até **29,64%**.
 - » Se usados quatro quase-identificadores, a chance de sucesso aumenta para **49,86%**.
 - » Com o uso de dez quase-identificadores, o risco é elevado para **75,51%**.
- » **Ataques ao Censo da Educação Superior 2018:**
 - » Nesse caso, o Inep destacou uma forma diferente de ataque, para tentar estimar a proporção de indivíduos que poderiam ser identificados a partir do cruzamento com outras bases contendo dados pessoais.
 - » Usando o *dia e ano de nascimento e código do curso* como quase-identificadores, os especialistas da UFMG afirmaram que foi possível

- reidentificar com absoluta certeza **39%** dos indivíduos, o que corresponde a aproximadamente 4,2 milhões de estudantes.
- › Usando dia, *mês* e ano de nascimento, além do código do curso, a porcentagem de indivíduos reidentificáveis subiu para **80%** (cerca de 8,6 milhões de estudantes).
 - › Considerando **11 atributos** como quase-identificadores, a reidentificação foi possível para **97,22%** dos indivíduos, o equivalente a 10.644.312 estudantes.

A partir desses resultados, o estudo da UFMG chegou à conclusão de que “a atual forma de divulgação dos censos educacionais pelo Inep submete os titulares dos dados a consideráveis riscos de violação de privacidade, incluindo reidentificação e inferência de atributos sensíveis” e que “tal situação poderia constituir violação da Lei Geral de Proteção de Dados Pessoais.”

CONTRAPONTO

Um dos grandes riscos de reidentificação é a presença da data de nascimento entre as variáveis da base de dados, como a UFMG apontou no Censo da Educação Superior 2018. No entanto, desde 2019 essa variável já havia sido retirada da base de dados, eliminando grande parte dos riscos anos antes da decisão de suspender a divulgação dos dados.

Já no caso do Censo da Educação Básica, a UFMG diz no estudo que não havia informação sobre a data completa de nascimento dos estudantes, apenas o mês e o ano de nascimento. Na mesma metodologia destacada pelo Inep sobre o Censo da Educação Superior, sem a data de nascimento seriam necessários usar 10 quase-identificadores para conseguir identificar **60,90%** dos indivíduos na base de dados, um valor bem abaixo do Censo da Educação Superior e usando a data de nascimento.

Isso indica que há quase-identificadores mais e menos eficazes em um ataque de reidentificação, mas que um dos principais – a data de nascimento – já não existia mais nas bases, o que já reduziu grande parte do risco à privacidade.

Ainda sobre data de nascimento, os pesquisadores que usam os dados na produção científica explicam que ela não é indispensável para fins de pesquisa, apenas a idade que a pessoa tinha na data de referência da coleta dos dados, para possibilitar cálculos como a distorção idade-série de grupos de estudantes.

2.2. Episódios em que o risco se concretizou

Em mais de um documento, o Inep reuniu diversos casos de vazamentos de dados pessoais envolvendo outros órgãos públicos, e três eventos em que houve “reidentificação” nas bases de dados do próprio órgão.

Na nota técnica publicada pela Deed³, há citação de seis notícias sobre falhas diferentes de proteção de dados educacionais:

1. Falha na Secretaria de Educação do Distrito Federal que expôs dados de estudantes
2. Falha em sistema do Ministério da Educação que expôs dados de servidores e até dos presidentes Bolsonaro e Lula
3. Vazamento de dados pessoais como e-mail, foto e senhas de milhares de estudantes de Pernambuco
4. Vazamento de dados da União Estadual de Estudantes de São Paulo que expôs dados de estudantes universitários
5. Falha em sistema de franquia de escolas que expôs dados de estudantes, funcionários e franqueados
6. Vazamento de dados da Secretaria Estadual de Educação de São Paulo que expôs dados de estudantes da rede

O Inep citou os casos acima afirmando que são “fatos que expõem continuamente a criticidade da questão” dos riscos aos dados pessoais.

CONTRAPONTO

Considerando a natureza dos microdados educacionais que o Inep divulgava publicamente até 2022, nenhuma das notícias acima pode ser considerada relacionada. Isso porque as notícias se referem a dados pessoais que nunca existiram nas versões públicas das bases divulgadas pelo Inep. Além disso, as falhas ocorreram em sistemas de acesso controlado, com a exposição indevida de dados pessoais, o que indica que os órgãos públicos precisam aprimorar a segurança da informação e os mecanismos de proteção desses dados, já que seguirão coletando informações pessoais e sensíveis de estudantes e profissionais da educação para a execução de políticas públicas.

Já as bases públicas de microdados educacionais do Inep estavam disponíveis para acesso de qualquer cidadão. Portanto, não se pode falar em risco de vazamento de dados que já estão livremente acessados.

O Inep, porém, foi além e citou três exemplos nos quais o risco de reidentificação individual se concretizou justamente nas suas bases públicas de microdados anonimizados.

Caso 1: Em novembro de 2015, dois pesquisadores publicaram artigo científico pela Universidade Federal da Paraíba (UFPB) que relata como a coautora da pesquisa foi capaz de “se encontrar” dentro dos microdados de professores do Censo da Educação Superior, ao buscar por uma combinação entre sua data de nascimento, seu sexo e o nome da instituição na qual ela trabalhava.⁴

Gustavo Motta, um dos dois coautores do estudo, afirmou que os pesquisadores nunca foram procurados pelo Inep para discutir os achados do levantamento.

A dupla, no entanto, escreveu para o órgão com a finalidade de alertar sobre a iminente publicação do artigo.

“Após a aceitação do artigo no SBSeg 2015 e antes de sua apresentação e publicação, encaminhamos em 8 de outubro de 2015 mensagem para o e-mail censosuperior@inep.gov.br do INEP informando que os ‘resultados do estudo mostraram que é possível identificar indivíduos de forma única, muito embora dados explícitos (como nome completo e, por ventura, números de documentos) tenham sido suprimidos da base de dados original, a fim de torná-la anônima, conforme determina a LAI’ e que o estudo seria apresentado/publicado no referido evento”, afirmou Motta.

O especialista disse ainda que considera adequada a decisão de “temporariamente não divulgar publicamente microdados individualizados”, e explicou que “não há técnica sistemática de anonimização capaz de resolver definitivamente o problema, mas sim de reduzir significativamente o risco de reidentificação, mas com custo elevado”.

Caso 2: Em abril de 2018, o jornal Folha de S.Paulo publicou uma reportagem com o título “Estudo inédito indica alta chance de fraude em mil provas do Enem”.⁵ Para produzir a reportagem, a equipe do jornal trabalhou em cima dos microdados do Exame Nacional do Ensino Médio (Enem), que trazem, para cada candidato, as respostas indicadas para cada questão de múltipla escolha.

A reportagem selecionou uma amostra de 10% das provas, representando as notas mais altas entre as edições 2011 e 2016, o que dá cerca de 3 milhões de provas do exame. E encontrou 1.125 provas com respostas tão semelhantes que era estatisticamente improvável serem parecidas por acaso, o que indica que pode ter havido fraude e cola entre candidatos.

A equipe então acessou informações públicas do resultado de processos seletivos usando a nota do Enem em universidades públicas, e conseguiu identificar, considerando a nota final divulgada pelas universidades, duas pessoas dentro do grupo de 1.125 provas suspeitas. Somente uma das pessoas foi contatada, e deu seu posicionamento na reportagem. A pessoa que não respondeu às demandas do jornal não teve o nome exposto na matéria.

CONTRAPONTO

Em ambos os casos, a reidentificação só foi possível a partir de informações auxiliares que não foram divulgadas pelo Inep.

Os dois pesquisadores da UFPB tinham posse de todos os dados pessoais necessários para “localizar” a professora na base de dados, que também era a titular das informações.

Já na reportagem, o jornalista de dados Daniel Mariani, responsável pelas análises que originaram a reportagem, afirmou que só foi possível identificar candidatos do Enem na base de dados do Inep porque a informação sobre o resultado de processos seletivos nas universidades públicas, como o vestibular de graduação, é pública. Além disso, segundo ele, o Inep jamais entrou em contato para pedir mais informações sobre como foi realizado o processo de reidentificação.

A Daeb, diretoria do Inep responsável pela realização do Enem e divulgação dos microdados do exame, chegou a confirmar, na nota técnica em que solicitou posicionamento oficial do Inep sobre a divulgação dos dados, que não buscou entender a fonte da reidentificação.

“Há evidências de que os participantes estão sendo identificados, ainda que não haja clareza quanto à razoabilidade dos meios utilizados”, diz a nota técnica. O documento também cita a LGPD que, no art. 5º, III, define o “dado anonimizado” como sendo “dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento”.

Sem saber se a identificação usou “meios técnicos razoáveis e disponíveis na ocasião de seu tratamento”, sequer é possível afirmar que tal identificação infringiu a LGPD. Além disso, vale ressaltar que a própria lei prevê fins legítimos para o tratamento de dados, inclusive o uso para atividades jornalísticas, como é o caso exposto pelo Inep: “Esta Lei não se aplica ao tratamento de dados pessoais (...) realizado para fins exclusivamente jornalístico e artísticos (Art. 4º, II, b).”

Caso 3: Em 2020, o Partido Novo lançou a candidatura de Filipe Sabará para a Prefeitura de São Paulo. A Revista Crusoé acessou o perfil do candidato no LinkedIn e procurou as instituições diretamente para verificar as informações divulgadas pelo próprio titular dos dados. Diante da não apresentação de documentos públicos, como o histórico escolar, comprovando a conclusão dos cursos, a revista publicou uma reportagem.⁶

Dois dias depois, o site O Antagonista acessou uma das bases de dados públicas do Censo da Educação Superior e procurou algum estudante que tenha a mesma data de nascimento de Filipe Sabará, e que tenha concluído o curso que ele disse ter feito no ano em que ele indicou a conclusão. A reportagem não encontrou nenhum indivíduo com essas três informações simultaneamente nos microdados. Ou seja, não houve reidentificação de fato, apenas não houve retorno positivo para os filtros usados na reportagem.⁷

CONTRAPONTO

Além de esse não ser um caso de reidentificação, pois nenhum indivíduo foi encontrado aplicando os filtros em questão, aqui também se trata mais uma vez de informações divulgadas publicamente pelo próprio indivíduo, que era candidato a cargo público e, portanto, sujeito à exposição de mais dados por meio das bases de dados de candidatos do Tribunal Superior Eleitoral (TSE), [8] inclusive dados mais sensíveis que formação acadêmica, como a declaração de bens, que, para os demais indivíduos, é protegida por sigilo fiscal.

2.3. Educação é alvo preferencial de ataques

Para o Inep, o setor educacional é um “alvo preferencial” de ataques cibernéticos. Na mesma nota da Deed de maio de 2021, a diretoria afirma que “um mapa de ameaças

cibernéticas, atualizado continuamente, elaborado por uma empresa de soluções de segurança da informação, apresenta a ‘indústria do setor educacional’ como um alvo preferencial de ataques (para mais informações consultar <https://threatmap.checkpoint.com>)”.

CONTRAPONTO

Uma visita ao site mostra que se trata de um mapa mundi com rastreamento em tempo real de ataques cibernéticos, com algumas informações consolidadas:

- » um *ranking* dos cinco países com o maior número de registros de ataques
- » um *ranking* das três “indústrias” mais visadas por ataques no último dia
- » um *ranking* dos três tipos de ataque mais frequentes

Para acompanhar o ranking em tempo real, foram realizadas visitas em cinco dias consecutivos, que mostraram que:

- » Os cinco países do ranking estão na África e na Ásia
- » A “indústria de educação” está no topo do ranking, com os órgãos governamentais se revezando na segunda e na terceira posições
- » Os três tipos de ataque cibernético mais comuns são:
 - › **phishing**: o envio de link pelo qual a pessoa é compelida a inserir seus próprios dados inadvertidamente, achando se tratar de um site oficial, e entregando seus dados de acesso a golpistas)
 - › **mobile**: acesso indevido a contas e perfis por meio de ataques a telefones celulares
 - › **worm**: um tipo de vírus que, ao se embrenhar no computador, consegue capturar informações

Ou seja, os alvos das ameaças no setor educacional provavelmente são instituições de ensino, e não o Inep. E nenhuma das três principais ameaças mapeadas pelo ThreatMap envolvem golpes relacionados a enriquecimento de dados ou uso de microdados anonimizados e já disponíveis publicamente, e sim a tentativas de invasão para acessar justamente os dados não anonimizados, um assunto não relacionado ao caso em questão.

2.4. Riscos subsequentes à segurança física dos indivíduos

Em 30 de dezembro de 2022, mais de dez meses depois de tirar os microdados do ar, o Inep concluiu a produção do Relatório de Impacto à Proteção de Dados Pessoais (RIPD),⁹ uma de quatro recomendações do Guia sobre Tratamento de Dados Pessoais pelo Poder Público publicado pela Autoridade Nacional de Proteção de Dados (ANPD), para que os órgãos tomem medidas de mitigação de risco para os titulares dos dados.

No documento, o Inep diz que “o impacto para a pessoa identificada reveste-se de características próprias, podendo alcançar consequências graves”, como atribuir às pessoas “e dados pessoais sensíveis como raça/cor, condição de saúde (deficiência e tipo), além de outras situações do seu vínculo escolar, como o rendimento escolar ao final do período (aprovação, reprovação ou abandono) e até a inferência sobre o seu desempenho ao longo do tempo”.

A autarquia argumenta ainda que “seria difícil de estimar” a extensão do impacto de alguém ter acesso aos detalhes acima de uma pessoa, pois seria possível saber o nível de escolaridade, a ocorrência de sucessos e insucessos na trajetória escolar, as movimentações no território, a expectativa de localização da pessoa e a abrangência de circulação.

A partir dessas informações, o Inep fornece alguns exemplos de “usos indevidos e incoerentes com a finalidade para a qual os dados pessoais foram originalmente tratados”, afirmando que a lista não é compreensiva. São eles:

- » **Perseguição e atentado contra a vítima**, por exemplo, um *stalker* poderia usar da informação para aumentar a sua chance de sucesso em caso de intentar alguma abordagem física da vítima.
- » **Enriquecimento de outras bases de dados e cadastros**, de domínio particular e/ou público para finalidades diversas, agregando-se a elas interesse e valor não compatíveis com a expectativa inicial daquelas bases de dados.
- » **Uso de informação pessoal para estelionato e golpes**, complementando informações requeridas em processos administrativos e/ou judiciais ou

servindo como validadores de competência do interlocutor junto ao titular em confirmação de situações pessoais.

- » **Aplicações em decisões automatizadas para excluir determinadas pessoas do quadro de clientes**, de um novo serviço, de processos seletivos diversos, de quadro de beneficiários, de processo de contratação de colaboradores, de campanhas publicitárias direcionadas.
- » **Perfilização** de hábitos, comportamentos, consumo, situação econômica, situação de saúde, preferências e interesses pessoais.

CONTRAPONTO

De acordo com especialistas, o enriquecimento de dados, ações de perseguição e atentados contra vítimas, além de submeter pessoas a tentativas de estelionato e golpes ou de provocar prejuízos como exclusão de um processo seletivo devido à cor/raça ou presença de deficiência são riscos reais. No entanto, são riscos cada vez mais disseminados e que afetam todos os órgãos públicos envolvidos na divulgação de dados oficiais, assim como todos os entes privados e as próprias pessoas que, individualmente, publicam informações sobre si mesmas. Nesses casos, inclusive, essas informações são muito mais valiosas para *stalkers*, posto que os microdados do Inep eram divulgados anualmente e com informações do ano anterior. Além disso, elas são muito menos sensíveis do que as informações de localização geográfica ou imagens compartilhadas por pessoas nas redes sociais, além de dados muito mais adequados para a perfilização de hábitos, comportamentos, consumo e interesses pessoais.

De acordo com os especialistas em privacidade de dados, é por isso que o texto da LGPD trabalha com o conceito de riscos e da atuação para mitigar os riscos, e recomendam que isso seja feito de forma ponderada. O RIPD do Inep, porém, traz uma abordagem considerada extrema das potenciais situações que poderiam ocorrer em caso de uma reidentificação.

2.5. Confiança na privacidade para garantir a qualidade dos dados

Em agosto de 2023, durante a XII Reunião da Associação Brasileira de Avaliação Educacional (Abave), em Campinas, a programação incluiu uma atividade sobre “Disponibilização de dados para pesquisa e monitoramento da educação: entre a transparência e a proteção dos estudantes”, que foi registrada em vídeo.⁹

Estiveram presentes Fábio Bravin, coordenador-geral de Controle de Qualidade e de Tratamento da Informação (CQTI) do Inep; Ernesto Martins Faria, especialista em cálculos e análises de indicadores educacionais e fundador do Interdisciplinaridade e Evidências no Debate Educacional (Iede); Clarissa Luz, advogada especialista em proteção de dados e sócia de Privacidade e Tecnologia do Felsberg Advogados; com moderação de Reynaldo Fernandes, professor de Economia na Universidade de São Paulo (USP) e ex-presidente do Inep entre 2005 e 2009, quando os microdados educacionais começaram a ser divulgados no formato que durou até 2022.

Ao longo do debate, Bravin abordou como a proteção da privacidade dos titulares dos dados (nesse caso, docentes e estudantes) era necessária não apenas para atender à LGPD, mas para garantir a qualidade final das bases de dados.

“O problema é: se eu tenho questionamentos acerca dos compromissos que nós deveríamos assumir em relação à pesquisa, a gente está fragilizando a própria pesquisa”, afirmou o coordenador-geral de CQTI do Inep. “A gente não quer colocar um conjunto de dados disponível, um microdado público, e daqui a pouco ter uma e petição falando assim: ‘Olha, ANPD, eu não concordo porque o Inep não está garantindo o meu direito’. E abrir uma petição contra o Inep e eu ter que tirar isso do ar de novo, e a gente ficar sem nada”, completou.

A consequência de pais deixarem de informar dados por desconfiar de potenciais riscos de identificação, segundo Bravin, seria o risco de queda na participação da coleta (que é censitária, ou seja, precisa abranger todos os alunos e professores), acarretando em muitos indivíduos na base de dados para os quais não há informação suficiente, reduzindo a qualidade e a representatividade final dos dados.

Para Bravin, em uma hipótese na qual as pessoas comecem a acreditar que o Censo “deixou de ser relevante porque ele não consegue mais retratar a realidade (...) do município A ou do município B, a gente vai fazer o quê com isso? Como é que a gente vai fazer pesquisa? Como é que a gente vai usar dados de referência oficial para reportar para os organismos internacionais?”.

ENTENDA MAIS

DADOS CENSITÁRIOS x DADOS AMOSTRAIS

Uma base de dados censitários, como o Censo Escolar e até o Censo Demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE), é uma base com informação sobre todos os indivíduos do universo em questão. Por exemplo, todos os estudantes de todas as escolas de Ensino Básico do Brasil, ou todos os professores trabalhando em escolas de Ensino Básico do Brasil.

Já os dados amostrais só trazem informações sobre alguns indivíduos desse universo. Nesses casos, pesquisadores aplicam técnicas estatísticas para definir quais indivíduos serão incluídos na amostra, para que ela seja “estatisticamente relevante”, ou seja, represente proporcionalmente o universo total. Um exemplo é a Pnad, Pesquisa Nacional por Amostra de Domicílios, realizada pelo IBGE.

Também, de acordo com ele, o contexto computacional mudou muito desde 2007, quando o Inep passou a coletar dados de todos os estudantes e professores – e a divulgar versões anonimizadas desses dados publicamente:

“Nós tínhamos outro cenário computacional, um cenário de conhecimento técnico, um cenário de tratamento de dados que não existiam naquela época.

Fazia sentido. (...) A própria LGPD vai dizer isso, que a anonimização é uma questão conjuntural. Se eu posso reverter [a anonimização], eu deixo de estar anonimizado. E isso, pode ser que, com o avanço da tecnologia, se torne um problema”, explicou o coordenador-geral.

CONTRAPONTO

Na mesma atividade da Abave, os demais especialistas participantes argumentaram, em resposta às falas de Fábio Bravin, que, embora os riscos sejam reais, a abordagem do problema é excessivamente extrema, especialmente considerando a extensão do efeito colateral de a sociedade brasileira já estar há três anos sem acesso aos microdados anonimizados.

Ainda em agosto de 2023, os especialistas explicaram que pouco se avançou objetivamente no estudo dos riscos reais, para que o Inep apresente uma estimativa mais detalhada da probabilidade de o risco se concretizar, para que haja uma tomada de posição mais clara. Houve a ressalva de que o estudo da UFMG também é limitado e foi realizado usando como base a data de nascimento, uma variável que já não estava mais disponível e que, sem ela, já reduzia em grande parte os riscos de reidentificação.

Eles afirmaram, também, que as soluções apresentadas até agora são muito limitadas e não conseguem dar conta das diversas demandas apresentadas pela sociedade, inclusive descentralizando o acesso dos dados aos órgãos estaduais e municipais de educação para que eles possam estender parcerias com pesquisadores.

Sobre a falta de confiança da população em fornecer os dados, o Inep, apesar de mencionar exemplos de falta de respondentes no caso do Censo Demográfico 2022 do IBGE, não chegou a apresentar estudo quantitativo ou qualitativo sobre a falta de confiança da comunidade nos censos educacionais. Além disso, os especialistas argumentaram que os pais fornecem os dados necessários para as escolas, e não diretamente ao Inep, ao contrário do que ocorre com o Censo Demográfico do IBGE, em que os recenseadores vão até a casa das pessoas para buscar informações, num

formato mais ostensivo e que pode gerar rejeição por receio da vigilância do Estado. Mas esse caso, segundo eles, não se aplica à questão de riscos de reidentificação em microdados anonimizados.

Por fim, eles disseram que, se o que falta é um posicionamento mais claro da ANPD sobre diversas questões que ainda não parecem definitivas o suficiente para oferecer segurança jurídica ao Inep, é possível agir ativamente produzindo mais artigos e estudos para auxiliar a acelerar esse processo.

2.6. O Sedap como alternativa à produção de pesquisa

Em 15 de março de 2022, a Câmara dos Deputados encaminhou ao Inep um ofício de autoria da deputada federal Marília Arraes requerendo “informações sobre a omissão de dados sobre o Exame Nacional do Ensino Médio (Enem) de 2020 e do Censo Escolar da Educação Básica de 2021” por parte da autarquia.

Entre os questionamentos feitos pelo Legislativo estava a seguinte pergunta: Qual seu impacto [da não divulgação dos microdados] no efetivo monitoramento dos índices educacionais por parte da população brasileira?”.

Em resposta, o Inep enviou à Câmara, em 1º de abril, um ofício¹⁰ argumentando o seguinte:

As pesquisas com a utilização dos dados tratados pelo Inep, eventualmente restringidos nos microdados, não estão inviabilizadas. O Inep possui o Serviço de Acesso a Dados Protegidos (Sedap), regido pela Portaria n. 637/2019 (referência 2.4), que utiliza protocolos de segurança, garantindo a manutenção do sigilo e da identidade dos indivíduos e instituições, conforme a legislação vigente. Pelo Sedap, o pesquisador poderá desenvolver estudos amplos e detalhados, considerando

tendências, padrões e trajetórias educacionais que podem ser traçados a partir de evidências apuradas pela Autarquia.

Atualmente, considerando a perspectiva de que os pedidos de acesso por meio do Sedap serão crescentes, o Instituto está analisando a viabilidade de implementação e operacionalização de outros meios capazes de ampliar o acesso aos microdados pelos pesquisadores, respeitando a legislação vigente de proteção de dados pessoais (ASPAR/GM/GM-MEC, Ofício nº 471/2022, p.2-3).

Essa não foi a primeira nem a última vez em que o Inep usou o Sedap e a sugestão de “outros meios” como substituição da divulgação pública dos microdados.

Em relação ao Sedap, a nota técnica da Deed de maio de 2021³ já recomendava a necessidade de “reforçar e ampliar a capacidade de atendimento do Sedap” e de “desenvolver estudos sobre a possibilidade de desenvolvimento de solução remota (não presencial) para o Sedap, ainda que com algumas limitações, para reduzir a pressão sobre o serviço presencial e como alternativa a usuários que poderiam ter suas necessidades atendidas pelo serviço”.

E no RIPD dos dados do Censo o Inep afirmou que “a necessidade de ampliação dos canais de acesso para pesquisadores já era observada e endereçada muito antes da alteração dos arquivos de microdados públicos”.¹¹

De fato, medidas foram tomadas para criar os núcleos Sedap já em 30 de março de 2022 e, como vimos no capítulo anterior, entre junho de 2023 e setembro de 2024 o Inep conseguiu firmar parcerias para instalar cinco salas remotas em três cidades (Belo Horizonte, João Pessoa e São Paulo).

Entre os “outros meios” que já tiveram algum tipo de detalhamento está a plataforma de acesso a dados, que foi anunciada em maio de 2024¹² e, segundo relatos feitos em audiências no Congresso Nacional, deveria ter começado a operar ainda no fim de 2024, mas ainda não estava no ar no fim de janeiro de 2025.

CONTRAPONTO

O Sedap, como vimos no capítulo anterior, foi criado em 2014 pelo Inep para dar acesso a dados protegidos, ou seja, às bases de dados contendo informações de fato pessoais dos indivíduos. Os microdados públicos eram anonimizados e não continham essas informações.

Além disso, a necessidade de estar fisicamente presente em um endereço para poder fazer a coleta de dados já impõe inúmeras restrições geográficas e financeiras, sem contar a disponibilidade de tempo, tanto para a parte da coleta de fato quanto para a produção da documentação previamente necessária para solicitar tal acesso.

Tal descentralização do Sedap em diversos locais físicos, com a possibilidade de expansão pela via remota, também pode aumentar a vulnerabilidade dos dados pessoais a ataques. Isso ocorre porque cada novo ponto na rede (como computadores e servidores) representa uma porta de entrada potencial para invasores. Assim, mais conexões significam mais oportunidades para exploração de vulnerabilidades.

Ainda, ela aumenta a dificuldade de monitoramento e controle: a tarefa de atualizar as permissões de acesso de todos os dispositivos e usuários se torna mais complexa, aumentando o risco de configurações incorretas ou permissões excessivas. Isso amplia a probabilidade de erro humano, pois cada ponto adicional pode ter um usuário que, se não estiver devidamente treinado em práticas de segurança, pode ser um elo fraco. Táticas como *phishing* e engenharia social, por exemplo, exploram falhas humanas.

Assim, ampliar o Sedap como resposta a acesso a dados que poderiam ser anonimizados e abertos não só não é uma solução adequada, como aumenta os riscos. Vale lembrar que a demanda por dados abertos é muito maior que a demanda por dados identificados.

Veremos em mais detalhes, no próximo capítulo, como se dá concretamente esse acesso, conforme relato de mais de uma pessoa que esteve no local. Também exploraremos como o Sedap e seus núcleos de acesso remoto foram incapazes de impedir a queda na produção científica usando os dados educacionais a partir de 2022.

2.7. Os censos educacionais são “pesquisas estatísticas” e não “registros administrativos”.

Adicionamos por último esse argumento porque, apesar de o Inep não embasar a decisão diretamente nele, é importante diferenciar os dois conceitos e entender por que o Inep sempre descreve os censos educacionais como “*pesquisas estatísticas*”.

A própria estrutura do site oficial do Inep abriga as páginas específicas sobre o Censo Escolar e o Censo da Educação Superior dentro da área “Pesquisas estatísticas e indicadores educacionais”.¹³

Em outro documento oficial, o Caderno de Conceitos e Orientações do Censo Escolar 2022,¹⁴ o Inep também descreve o Censo Escolar como “*uma pesquisa estatística declaratória realizada anualmente*” pelo Inep “em regime de colaboração entre a União, os estados, o Distrito Federal e os municípios”.

Nessa cartilha, o instituto diz que o Censo “tem por objetivo realizar um amplo levantamento sobre a educação brasileira” e que ele “é o mais importante levantamento estatístico educacional sobre as diferentes etapas e modalidades de ensino da educação básica e da educação profissional”.

ENTENDA MAIS

PESQUISAS ESTATÍSTICAS

A pesquisa estatística é uma produção de informações com a função de fornecer à população como um todo um conjunto de informações necessárias à população e ao Estado. Ela é regida por princípios básicos de qualidade que têm que ser seguidos, desde a forma como é coletada a informação ao fato de que você pode publicar o erro associado à pesquisa, ou seja, os métodos permitem calcular o erro e os intervalos de confiança, que auxiliam os usuários a terem uma ideia da qualidade daquela estatística.

As estatísticas podem ser oficiais, produzidas por órgãos públicos, ou não, sendo realizadas por entidades privadas.

No caso das estatísticas oficiais, elas devem seguir padrões de qualidade porque também são usadas oficialmente pelo país em âmbitos internacionais.

Mas há duas informações importantes, nesse mesmo documento, que indicam que o Censo Escolar pode também ser considerado um **registro administrativo**.

Na apresentação que faz aos gestores escolares responsáveis pelo preenchimento do Censo, o Inep cita duas leis que determinam que o Censo Escolar deve ser usado como informação oficial para a execução de políticas públicas.

A primeira delas é o Plano Nacional de Educação (PNE), que impôs metas aos diversos entes governamentais e prazos para que fossem cumpridas:

Em 2014, foi aprovado o **Plano Nacional de Educação (PNE)** – Lei nº 13.005, de 25 de junho –, que **atribui ao Inep a realização de estudos e pesquisas das metas do Plano, tendo como uma das referências os censos educacionais nacionais**.

O Censo Escolar é, portanto, ferramenta indispensável para que os diversos atores educacionais e a sociedade em geral possam conhecer a situação educacional do País, do Distrito Federal, dos estados e dos municípios, de modo a acompanhar a efetividade das políticas públicas (INEP, Caderno de Conceitos e Orientações do Censo Escolar 2022, p.2, grifo nosso).

A segunda lei é mais recente, a aprovação do **Novo Fundo de Manutenção e Desenvolvimento da Educação Básica (Fundeb)**, o principal instrumento para financiar as escolas públicas, com base na quantidade de matrículas. Diz o Inep:

Mais recentemente foi normatizado o novo Fundo de Manutenção e Desenvolvimento da Educação Básica – Fundeb (Lei nº 14.113, de 25 de dezembro de 2020, Lei nº 14.276, de 27 de dezembro de 2021 e Decreto nº 10.656, de 22 de

março de 2021), o qual **determina que, para fins de distribuição dos recursos do Fundo, serão consideradas as matrículas presenciais efetivas apuradas no Censo Escolar mais atualizado**. Além disso, indica que os estados, o Distrito Federal e os municípios terão 30 dias, contados da publicação dos dados preliminares do Censo Escolar, para ratificar ou retificar os dados publicados, sob pena de responsabilização administrativa. Enfatiza ainda que, após a sua publicação final, as informações do Censo Escolar da Educação Básica comporão as estatísticas oficiais da educação básica, vedada qualquer alteração dos dados (INEP, Caderno de Conceitos e Orientações do Censo Escolar 2022, p.2, grifo nosso).

ENTENDA MAIS

REGISTRO ADMINISTRATIVO

A definição mais direta de um registro administrativo é um conjunto de **dados que são levantados para a administração do Estado**. Ou seja, suas informações são necessárias de uma forma operacional, para fornecer informações necessárias para a gestão pública. Por exemplo: a declaração de **imposto de renda de pessoa física ou jurídica** é necessária para a Receita Federal saber quanto deve cobrar de imposto. Uma **nota fiscal eletrônica** também é um registro administrativo. As bases de dados da **Relação Anual de Informações Sociais (Rais)** e o **Cadastro Geral de Empregados e Desempregados (Caged)** são registros administrativos segundo o próprio Ministério do Trabalho e Emprego (MTE), responsável por sua elaboração e divulgação – também em formato de microdados anonimizados, contendo inclusive informações sensíveis como o salário de cada empregado, no caso da Rais.

De acordo com o MTE, as duas bases são “ferramentas fundamentais para a implementação dos Programas Governamentais integrantes do Sistema Público de Emprego, para elaboração de Políticas Públicas e para estudos e pesquisas balizadores das Ações de Emprego”.¹⁵

Além de serem informações coletadas com o princípio primordial de garantir a administração do Estado, outra característica do registro administrativo é uma **flexibilidade maior no rigor estatístico**. A divulgação dos registros não necessariamente precisa passar por um tratamento e correção de dados ou ser acompanhada de detalhes sobre a margem de erro ou o intervalo de confiança dos dados.

Não há normas específicas ditando quais registros devem ser divulgados publicamente, nem quando ou como. Mas, na **Constituição Federal**,¹⁶ há alguns artigos sobre o assunto:

- » **Art. 5º XIV:** é assegurado a todos o acesso à informação e resguardado o sigilo da fonte, quando necessário ao exercício profissional.
- » **Art. 5º XXXIII:** todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado.
- » **Art. 37:** A administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios obedecerá aos princípios de legalidade, impessoalidade, moralidade, publicidade e, também, ao seguinte:
 - » (...) § 3º: A lei disciplinará as formas de participação do usuário na administração pública direta e indireta, regulando especialmente:
 - (...) II: o acesso dos usuários a **registros administrativos** e a informações sobre atos de governo, observado o disposto no art. 5º, X e XXXIII.
- » **Art. 216. § 2º:** Cabem à administração pública, na forma da lei, a gestão da documentação governamental e as providências para franquear sua consulta a quantos dela necessitem.

Ainda assim, eles podem ser usados como estatística. Para que isso aconteça, é preciso aplicar as metodologias de controle de qualidade que a mesma apresenta.

Ambos os tipos de documento – estatísticas ou registros administrativos – devem se submeter às normas de sigilo e proteção de dados, ou seja, estão proibidos de divulgar publicamente dados pessoais, resguardadas algumas exceções, seguindo o que determina a legislação.

ENTENDA MAIS

TIPOS DE SIGILO

No caso de **sigilo fiscal** (acesso a dados sobre renda), do **sigilo bancário** (acesso às movimentações financeiras) ou do **sigilo telemático** (acesso às chamadas telefônicas ou aplicativos de celular), apesar da proteção da intimidade garantida pela Constituição Federal, é possível que ela seja quebrada por decisão judicial, a pedido de autoridades durante uma investigação.

Mas o **sigilo estatístico**, por sua vez, está amparado por lei há quase 60 anos – especificamente o do IBGE: “A atuação do IBGE está amparada em legislação federal específica, a da Lei nº 5.534/1968, conhecida como Lei do Sigilo Estatístico, e também na prática da maioria dos países e nas recomendações do Instituto Internacional de Estatística”,¹⁷ diz o órgão.

A proteção foi reafirmada em 2017 pelo STF.¹⁸ No entanto, o próprio IBGE, por sua vez, teve uma derrota no Supremo em outro caso, durante a pandemia. Devido ao isolamento social, o governo federal publicou, a pedido do órgão, uma medida provisória determinando que as companhias de telefonia cedessem ao IBGE o nome, número de telefone e endereço de seus clientes, para que o instituto pudesse manter as pesquisas domiciliares de forma remota.

A MP 954/2020 foi publicada em 17 de abril de 2020,¹⁹ mas foi suspensa por medida cautelar pelo STF em 25 de abril, por considerar que o compartilhamento do cadastro implicava na violação de dados pessoais.²⁰ Em 7 de maio, o plenário confirmou a suspensão da eficácia,²¹ e ela perdeu sua validade em agosto daquele ano.²²

No entanto, não existem atualmente normas muito detalhadas sobre como enquadrar as diferentes bases de dados nas duas classificações, ou quais regras detalhadas cada classificação dessas deve seguir em termos de como tratar os dados.

Até mesmo o Inep levantou essa incerteza jurídica em 2021. Na nota técnica produzida pela Diretoria de Avaliação da Educação Básica (Daeb) em julho desse ano,²³ a respeito dos microdados do Enem, a Coordenação-Geral de Instrumentos e Medidas ressalta que “é importante confirmar se o Inep será caracterizado como órgão de pesquisa” segundo o conceito de “órgão de pesquisa” definido pela LGPD,²⁴ que consiste em:

1. ser um “órgão ou entidade da administração pública direta ou indireta ou pessoa jurídica de direito privado, sem fins lucrativos legalmente constituída sob as leis brasileiras”;
2. ter “sede e foro no país”;
3. incluir “em sua missão institucional ou em seu objetivo social ou estatutário a pesquisa básica ou aplicada de caráter histórico, científico, tecnológico ou estatístico”.

Diz a Cgim/Daeb:

Parece evidente que o Inep atende plenamente os dois primeiros requisitos legais. Entretanto, o termo ‘Pesquisa’ não aparece diretamente na missão institucional do Inep. Contudo, aparece o termo ‘Produzir conhecimento’, implicitamente associado à realização de pesquisas. É preciso, portanto, definir com clareza se o Inep se enquadra na categoria de órgão de pesquisa prevista na legislação, para que possa usufruir das prerrogativas, apesar de não cumprir, pelo menos explicitamente, um dos requisitos previstos na legislação (CGIM/DAEB/INEP, Nota Técnica 14/2021, p.4).

Por que a definição importa? De acordo com o próprio Inep, caso seja reconhecido como um órgão de pesquisa, ele passa a se enquadrar em um dos nove casos em que o órgão não necessita que cada titular dos dados assine um termo de consentimento para o tratamento dos dados. Três dessas hipóteses de uso dos dados são citadas na nota técnica para afirmar que ainda é necessária a confirmação sobre o enquadramento do Inep:

1. para cumprir “obrigação legal ou regulatória pelo controlador”;

2. para executar “políticas públicas previstas em leis e regulamentos ou respaldadas em contratos, convênios ou instrumentos congêneres”;
3. para a “realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais”.

2.8. Conclusão

Como vimos acima, há vários argumentos diferentes levantados pelo Inep para justificar a mudança da política de divulgação dos microdados educacionais. E, embora os especialistas concordem com a existência dos riscos, a principal discordância está mais focada nas propostas de solução apresentadas até agora, consideradas extremas demais, empurrando o pêndulo para muito longe dos avanços de transparência conquistados após a Lei de Acesso à Informação (LAI). Avanços que, conforme os próprios órgãos de controle, e a própria LGPD, podem coexistir em equilíbrio com a proteção de dados.

No próximo capítulo, vamos ver mais a fundo como a mudança foi recebida pelos diversos setores da sociedade e poder público, além do impacto que ela teve ao longo dos últimos três anos.

Referências

1. Disponível em: <<https://www.gov.br/Inep/pt-br/assuntos/noticias/institucional/nota-de-esclarecimento-divulgacao-dos-microdados>>.
2. Disponível em: <https://download.inep.gov.br/microdados/TED_8750-UFGM.pdf>.
3. Disponível em: <https://download.inep.gov.br/microdados/nota_tecnica_5-2021_deed.pdf>.
4. Disponível em: <<https://sol.sbc.org.br/indINSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA NOTA TÉCNICA Nº 5/2021/CGCQTI/DEED PROCESSO Nº 23036.ex.php/sbseg/article/view/20116>>.
5. Disponível em: <<https://www1.folha.uol.com.br/educacao/2018/04/estudo-inedito-indica-alta-chance-de-fraude-em-mil-provas-do-enem.shtml>>.
6. Disponível em: <https://crusoe.com.br/diario/a-la-decotelli-inconsistencias-em-curriculo-derrubam-candidato-do-novo-em-sp/?utm_source=crs-site&utm_medium=crs-login&utm_campaign=redir>.
7. Disponível em: <<https://oantagonista.com.br/brasil/dados-de-filipe-sabara-nao-aparecem-em-censo-do-mec/>>.
8. Disponível em: <<https://divulgacandcontas.tse.jus.br/divulga/#/home>>.
9. Disponível em: <<https://youtu.be/b61HeLzasRs?feature=shared&t=4920>>.
10. Disponível em: <https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2155098> (p.2-3)
11. Disponível em: <https://download.inep.gov.br/microdados/RIPD_dos_microdados_dos_Censos_da_Educa%C3%A7%C3%A3o.pdf>.
12. Disponível em: <<https://www.gov.br/Inep/pt-br/assuntos/noticias/institucional/Inep-apresenta-nova-plataforma-de-dados>>.
13. Disponível em: <<https://www.gov.br/Inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar>>.
14. Disponível em: <https://download.inep.gov.br/pesquisas_estatisticas_indicadores_educacionais/censo_escolar/orientacoes/matrricula_inicial/caderno_de_conceitos_e_orientacoes_censo_escolar_2022.pdf>.
15. Disponível em: <<https://www.gov.br/trabalho-e-emprego/pt-br/assuntos/estatisticas-trabalho/o-pdet/registros-administrativos>>.
16. Disponível em: <https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>.
17. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/33288-protexao-e-confidencialidade-dos-dados-sao-compromisso-permanente-do-ibge>>.
18. Disponível em: <<https://www.conjur.com.br/2017-mai-05/carmen-derruba-ordem-ibge-repassar-dados-criancas-mpf>>.
19. Disponível em: <<https://www12.senado.leg.br/publicacoes/estudos-legislativos/tipos-de-estudos/sumarios-de-proposicoes/mpv954>>.
20. Disponível em: <<https://www.ibge.gov.br/novo-portal-destaques/27493-comunicado-sobre-a-mp-954-2020-decisao-do-stf.html>>.
21. Disponível em: <<https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=442902>>.
22. Disponível em: <<https://www12.senado.leg.br/noticias/materias/2020/08/17/perde-eficacia-mp-que-obrigava-operadoras-a-compartilhar-dados-com-o-ibge>>.
23. Disponível em: <https://download.inep.gov.br/microdados/nota_tecnica_14-2021_daeb.pdf> (p.4).
24. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm>.

3. REPERCUSSÃO

A mudança na política do Inep para a divulgação de microdados educacionais teve impactos profundos em diversos setores da sociedade. Das organizações atuantes na área de Transparência e Privacidade aos tribunais de conta, passando pelos pesquisadores nas áreas de Educação, Economia e outras, e inclusive no Poder Legislativo, a retirada das bases de dados do ar foi veementemente criticada, em especial por ter sido feita sem aviso prévio e diálogo com os usuários dos dados.

Neste capítulo, veremos como a nova política repercutiu entre esses setores, que impacto ela teve no acesso à informação e produção acadêmica e as iniciativas motivadas pela mudança.

3.1. Sociedade civil

» Posicionamento

Os primeiros pronunciamentos da sociedade civil organizada começaram a sair quatro dias após a retirada do ar das bases de dados do Inep. Em 22 de fevereiro de 2022, duas notas foram publicadas: uma conjunta, reunindo 36 organizações das áreas de Pesquisa em Educação, como a Campanha Nacional pelo Direito à Educação (CNDE) e a Associação Nacional de Pesquisa em Financiamento da Educação (Fineduca); transparência, incluindo a Open Knowledge Brasil (OKBR); formação de professores, a exemplo do Fórum Nacional de Formação de Professores da Educação Básica (FORPARFOR); e defesa das crianças e adolescentes e minorias, como o Centro de Defesa da Criança e do Adolescente do Ceará (Cedeca-Ceará) e a Coordenação Nacional de Articulação de Quilombos (Conaq).¹

Juntas, as entidades registraram sua preocupação com o uso da LGPD “como justificativa genérica para o descarte dos microdados do Censo Escolar” sem apresentar o fundamento legal. “É preciso proteger a privacidade, sem abdicar da transparência”, afirmaram as organizações, ressaltando que “a própria LGPD deixa claro em seu artigo 7º, incisos II e III, que a administração pública pode realizar o tratamento de dados pessoais necessários ao cumprimento de obrigação legal

e/ou execução de políticas públicas, sem que para isso seja necessário o prévio consentimento da/o titular destes dados”.

Na opinião das 36 entidades:

A medida do Inep de reduzir drasticamente o detalhamento dos dados da última edição do Censo, além de remover todas as bases históricas, é desproporcional e não está alinhada às melhores práticas internacionais de avaliação de risco e impacto de publicação de dados previstas na própria LGPD. Como exposto, os benefícios decorrentes do uso dos microdados por diversos setores da sociedade vêm tendo seu valor comprovado há mais de uma década e estão amplamente disponíveis para avaliação. Eventuais riscos de reidentificação, por outro lado, tendem a ser isolados e seu impacto não foi demonstrado pelo órgão, já que estudos e pareceres técnicos não foram tornados públicos. Uma política séria de proteção de dados também precisa ser transparente e garantir a preponderância do interesse público (CNDE et al. Posicionamento Público sobre Microdados do Enem e do Censo Escolar, 2022, p.2).

A segunda nota de 22 de fevereiro foi publicada pelo ***Interdisciplinaridade e Evidências no Debate Educacional (Iede)***², que listou alternativas para aumentar a proteção dos dados sem a necessidade de retirar por completo os microdados do ar:

Sem os microdados do Censo Escolar, os pesquisadores não conseguem, por exemplo, fazer análises demográficas de acesso educacional por etapa de ensino ou por localização da escola (urbanas versus escolas rurais). São inviáveis ainda análises por cor/raça dos estudantes ou ainda considerando características específicas deles, como deficiência ou superdotação. No caso dos professores, faltam também informações sobre cor/raça, idade, formação e vínculo com a escola. Esses dados são essenciais para se conhecer melhor quem são os profissionais que atuam na Educação Básica para que, a partir disso, sejam criadas políticas que atendam às necessidades deles (IEDE, Nota Técnica sobre Microdados do Enem e do Censo Escolar, 2022).

No dia seguinte, em 23 de fevereiro, o ***Fórum de Direito de Acesso a Informações Públicas*** também divulgou nota contrária à retirada do ar dos microdados.³

Composto na época por 25 organizações da sociedade civil, entre elas a Associação de Jornalistas de Educação (Jeduca), a Associação Brasileira de Jornalismo Investigativo (Abraji), a Associação Fiquem Sabendo e a Open

Knowledge Brasil, o fórum opinou que a decisão do Inep de suprimir dados antes divulgados publicamente “é equivocada e compromete a transparência das políticas públicas de educação”. O grupo também mencionou exemplos concretos de como os dados já contribuíram para o debate sobre a melhoria da qualidade da educação brasileira:

“Reportagens usaram os dados para mostrar que só 293 estudantes nas piores condições socioeconômicas obtiveram notas semelhantes a de alunos de escolas de elite no Enem, e para verificar o desempenho de alunos cotistas e não cotistas em universidades, por exemplo. Iniciativas desse tipo já foram reconhecidas pelo próprio Inep através de seu Prêmio de Jornalismo. A partir de dados do Censo Escolar, pesquisadores produziram um Indicador de Permanência Escolar, que fornece um panorama de quantos jovens chegaram a ingressar na escola, mas em algum momento abandonaram os estudos.” (Fórum de Direito de Acesso a Informações Públicas)

A nota ressaltou ainda que “simplesmente retirar as informações é desproporcional e fere a própria LGPD” e destacou que “os benefícios dos dados já anonimizados estão colocados há mais de uma década, e geram imenso valor para a sociedade, como explicado aqui. Já os riscos, se existem, são exceções e casos isolados, que não podem se sobrepor ao interesse geral”.

» Produção de conhecimento

Em 10 de março, o Iede divulgou um novo estudo,⁴ apresentando uma análise empírica de outras formas de manter a divulgação dos microdados educacionais, mas reduzindo os riscos de reidentificação de indivíduos apresentados no estudo da UFMG.

Com o objetivo de mostrar que “mesmo uma interpretação extrema – e equivocada – da LGPD não inviabilizaria a divulgação dos microdados”, já que “há muitas opções, a partir do ajuste em variáveis, de reduzir drasticamente ou mesmo zerar a possibilidade de re-identificação dos titulares dos dados”, o Iede propôs um exercício com a base de dados do Censo Escolar de 2020.

A entidade explica que, “com a realização de ajustes em algumas variáveis e a exclusão de (poucas) outras, é possível pseudoanonimizar os dados de todos os

estudantes”, ressaltando, que, segundo a própria LGPD, a pseudonimização é o “tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro”.

No estudo, a área de Pesquisa do Iede separou as variáveis das bases de dados em três grupos:

- » variáveis que poderiam ser **adaptadas**;
- » variáveis que poderiam ser no máximo **mascaradas** (mas devem permanecer na base porque são consideradas fundamentais à pesquisa);
- » variáveis que poderiam ser **excluídas** (porque facilitam a identificação dos estudantes, mas não são as imprescindíveis à maioria das pesquisas).

Sobre as variáveis relativas à **data de nascimento** dos indivíduos, o exercício propôs indicar apenas a idade da pessoa em abril e em julho, já que, pelas regras do Ensino Fundamental, o ingresso dos alunos nessa etapa de ensino leva em conta a idade em 31 de março e, para efeitos de estimativas populacionais, a data levada em conta é 1º de julho.

A variável relativa à **nacionalidade** foi agrupada, unindo os estudantes brasileiros e os que são brasileiros, mas nasceram no exterior. Assim, a variável passa a ter apenas duas informações: se o indivíduo é brasileiro e se é estrangeiro. Como há poucas pessoas brasileiras nascidas no exterior, a probabilidade de elas serem identificadas nas bases por meio desse filtro era alta, mas a sugestão elimina esse risco.

A variável sobre o **local onde o estudante recebe as aulas** também sofreu agrupamento. Antes, era possível saber quem tinha aulas em hospital e em domicílio separadamente. Com a sugestão, esse detalhamento fica menor, entretanto reduz os riscos de reidentificação.

O estudo sugeriu ainda a exclusão das variáveis de **mês de nascimento** e **município** e **país de nascimento** do indivíduo.

Já para as variáveis com dados dos *alunos*, *escola* e *município da escola*, como a coluna “ID_ALUNO”, a sugestão foi criar um código aleatório que seja o mesmo nas diversas bases, para possibilitar estudos longitudinais sem que seja possível ao público acessar diretamente essas informações.

Uma última sugestão feita pelo lede é *agrupar indivíduos que sejam considerados outliers* dentro da base de dados. Por exemplo: estudantes mais velhos existem em pouca quantidade e podem ser identificados mais facilmente. A solução, aqui, seria agrupar estudantes a partir de uma certa idade, o que reduziria os detalhes sobre essa pequena população, mas sem comprometer estatisticamente o caráter censitário das bases de dados.

“Após todos esses cuidados, de um universo de 3,7 milhões de alunos, chegou-se a apenas 224 matriculados que possuem ‘informações únicas’ em suas mesorregiões, o que representa 0,006% do total”, explicou o lede no estudo.

Já a *Open Knowledge Brasil* e a *Data Privacy Brasil* realizaram, em 29 de abril, uma atividade fechada⁵ para facilitar o diálogo qualificado entre as diferentes perspectivas dos atores envolvidos com o tema a respeito da operacionalização de uma política de dados abertos que assegure, simultaneamente, o direito à proteção de dados pessoais dos cidadãos.

Um relatório elaborado a partir da atividade foi publicado em setembro de 2022,⁶ com o objetivo de “orientar pesquisas e tomadas de decisão em casos similares, de modo que a *proteção de dados pessoais e transparência pública caminhem em harmonia*”.

O documento enfatiza a necessidade de que o processo de análise dos riscos e benefícios do tratamento de dados pessoais seja participativo e de escuta da sociedade, e ressalta que a discussão a respeito de divulgar os dados não pode ser baseada apenas em critérios técnicos. É preciso levar em conta a análise qualitativa por meio de um debate público e democrático.

» Promoção de debate público

Em abril, a Associação de Jornalistas de Educação (Jeduca) promoveu dois *webinários*⁷ reunindo representantes do Inep, do estudo realizado pela UFMG,

do jornalismo educacional, da pesquisa na área de Educação e especialistas em Transparência para debater a mudança na política.

Cada webinar teve uma hora e meia de duração. O primeiro, realizado em 12 de abril, focou nos **aspectos técnicos**.⁸

O Inep e a UFMG apresentaram as justificativas para a decisão tomada em fevereiro (exploradas no capítulo anterior). Em contrapartida, representantes do jornalismo e da pesquisa em educação ressaltaram, segundo a Jeduca, “o interesse público dos dados e informações produzidas pelo Inep, especialmente para o enfrentamento das desigualdades educacionais, já que eles possibilitam um olhar no detalhe, favorecendo a formulação de programas alinhados com as necessidades e características específicas”.

A sociedade civil ainda propôs alternativas para manter os microdados públicos, como assinatura de compromisso pelas pessoas que tiverem acesso ao material, mascaramento de parte das informações, entre outras.

“Também foi feito um alerta quanto à necessidade de se implementar rapidamente meios de acesso aos microdados, já que muitas pesquisas acadêmicas foram inviabilizadas com a suspensão da divulgação pública.”

Já no segundo webinar, em 14 de abril, o foco ficou nos **impactos e aspectos legais**.⁹ Os participantes da sociedade civil ressaltaram que a medida foi tomada sem a realização de uma avaliação de risco, “um instrumento previsto pela LGPD, através do qual é possível estabelecer parâmetros e consensos relativos à divulgação de informações pessoais”.

Além disso, os participantes criticaram a ausência de diálogo do Inep com a sociedade civil, “já que poderiam ter sido pensadas em conjunto estratégias para divulgar os microdados e, ao mesmo tempo, preservar o sigilo das informações pessoais”.

Nos anos seguintes à mudança da política, a mobilização em prol da revisão da medida se manteve. Em maio de 2023, dez especialistas e representantes de

entidades da sociedade civil sem fins lucrativos e empresas destinadas à produção de pesquisa, como Instituto Natural, Comunidade Educativa Cedac (atual Roda Educativa), Iede e Fundação Roberto Marinho, coassinaram um ofício ao Inep solicitando acesso aos microdados.

No documento, ao qual a equipe da OKBR teve acesso, as entidades afirmam que os estudos realizados por elas “são diferentes entre si, mas todos têm um objetivo em comum: a busca pelo aprimoramento de suas atuações a fim de contribuir para melhorar a educação básica pública do país. Para isso, são desenvolvidas pesquisas para monitorar e avaliar a qualidade da educação no país, compreender desafios, auxiliar escolas e redes de ensino na superação dos mesmos e identificar casos de sucesso que possam inspirar seus pares, entre outros”.

É ressaltado, porém, nesse texto que os novos formatos de bases de dados divulgados pelo Inep não permitem que sejam “analisados de forma adequada”, pois ocultam informações essenciais para a realização dos estudos.

Por fim, as entidades se colocam disponíveis para dialogar com o Inep e solicitam que o órgão aceite “atender a solicitações específicas, com objetivos declarados e firmando um compromisso de manutenção do sigilo das informações individuais pelos solicitantes”.

Segundo pessoas signatárias da carta, o Inep não atendeu à solicitação. Meses depois, em agosto, a **XII Reunião da Abave** dedicou uma mesa ao tema, com a presença de Fábio Bravin, do Inep; Ernesto Martins Faria, do Iede; Clarissa Luz, advogada especialista na Lei Geral de Proteção de Dados; e Reynaldo Fernandes, ex-presidente do Inep e então vice-presidente da Abave.¹⁰

No evento realizado em Campinas, como vimos em detalhes no capítulo anterior, a comunidade de pesquisa acadêmica na área de Educação argumentou com o Inep a respeito da leitura considerada extremista da LGPD, além de cobrar maior celeridade para oferecer uma alternativa de acesso aos microdados. Naquele momento, as bases de dados no formato anterior já estavam indisponíveis havia um ano e meio.

Ainda em 2023, em outubro, a Associação Nacional de Pós-graduação e Pesquisa em Educação (Anped) aprovou uma moção, durante sua 41ª Reunião Nacional,¹¹ solicitando que o Inep revise sua nova política de não divulgar os microdados.

No documento, a Anped afirma que “a sua divulgação [dos dados] apenas no nível de escola ou o acesso restrito – por meio exclusivamente da solicitação junto ao Serviço de Acesso a Dados Protegidos (Sedap) – dificultam a democratização de informações relevantes para o trabalho acadêmico e o acompanhamento social”, e solicita que “a necessária proteção aos dados pessoais não limite o desenvolvimento de pesquisas que visam a garantia do direito à educação, na forma do Art. 205 da Constituição Federal”.

Em dezembro do mesmo ano, a Fineduca (Associação Nacional de Pesquisa em Financiamento da Educação), durante seu XI Encontro Anual, aprovou, entre as medidas recomendadas pela entidade, “a retomada de disponibilização pública dos dados educacionais do censo da educação básica e da educação superior e das avaliações, no menor nível de desagregação”.¹²

3.2. Controle externo

Desde a mudança na política do Inep, os tribunais de contas estaduais têm se mostrado uma das principais vozes pela retomada da divulgação dos microdados educacionais.

Além de questões relativas à transparência ativa de informações, essa postura se deve em grande parte à adoção de metodologias baseadas em evidências na fiscalização do trabalho das prefeituras em todo o Brasil. Por meio de organizações como a Associação dos Membros dos Tribunais de Contas do Brasil (Atricon) e o Instituto Rui Barbosa (IRB), os tribunais fizeram enormes avanços no uso dos dados educacionais do Inep para acompanhar o trabalho dos gestores municipais.

O trabalho se tornou ainda mais relevante após a aprovação do Plano Nacional da Educação (PNE) 2014-2024 pelo Congresso Nacional.

A Atricon e o Instituto Rui Barbosa não usaram os dados apenas em atuação interna, mas também lançaram o TC-Educa, uma plataforma pública de monitoramento de metas do PNE, com foco inicial na expansão de matrículas na Educação Infantil, disponível a toda sociedade.

Ao longo do tempo, a ação de fiscalização dos tribunais de conta passou a ser dirigida por dados também na área de Infraestrutura das escolas, como acesso à internet, e em outros indicadores, a exemplo de taxas de evasão e reprovação. Ao todo, até 2020 os dados eram usados para a fiscalização de mais de dez metas do PNE.

Para as análises dos órgãos de controle externo, bastava o acesso aos microdados públicos do Censo Escolar no formato anterior, que dispunham de informações no nível dos estudantes e professores. Mas, ao retirar essas bases do ar, o Inep acabou inviabilizando boa parte do trabalho que subsidia a fiscalização dos prefeitos brasileiros.

Três anos após essa mudança, os tribunais de conta ainda não contam com nenhum tipo de acesso direto a essas bases.

Como primeira reação, uma nota conjunta publicada em 30 de março de 2022¹³ foi assinada de forma **unânime** pelas entidades que compõem o Sistema Tribunais de Contas: Atricon, IRB, Associação Brasileira dos Tribunais de Contas dos Municípios (Abracom), Conselho Nacional de Presidentes dos Tribunais de Contas (CNPTC), Associação Nacional dos Ministros e Conselheiros Substitutos dos Tribunais de Contas (Audicon), Associação Nacional do Ministério Público de Contas (Ampcon), Conselho Nacional de Procuradores-Gerais de Contas (CNPGC) e Associação Nacional dos Auditores de Controle Externo dos Tribunais de Contas do Brasil (ANTC).

As entidades afirmam que “a posição do Inep dificulta que organismos públicos, entidades da sociedade civil e os próprios cidadãos monitorem o cumprimento do Plano Nacional de Educação, bem como que atuem no sentido de reduzir as desigualdades sob perspectivas de raça, gênero, nível socioeconômico, dentre outras análises” e que é “necessário agir de forma proporcional para, ao mesmo tempo, garantir a segurança das informações, diminuir os riscos de identificação pessoal e viabilizar o acesso aos dados necessários para pesquisas em educação”.¹⁴

Segundo Cezar Miola, atual presidente da Atricon e conselheiro do Tribunal de Contas do Estado do Rio Grande do Sul (TCE-RS), “ao longo dos últimos anos, pouquíssimas manifestações no âmbito das entidades do sistema de controle externo reuniram tamanho apoio”.

O conselheiro explica que os microdados forneciam aos tribunais de conta diversas vantagens. Uma delas é a otimização de recursos. “Os microdados vinham servindo de elementos para definir **estratégias de fiscalização** do tribunal de contas”. Para ele, “com os dados do Censo Escolar e do IBGE, é possível definir um olhar do controle externo por hipótese, identificar os locais com atendimento mais crítico”.

A outra vantagem é aumentar a **credibilidade das fiscalizações**, libertando os tribunais de contas de uma atuação muito formal em cima de contratos e licitações. “Quando fomos dar um verdadeiro salto de natureza qualitativa temos essa limitação”, afirma o presidente da Atricon, ressaltando que a falta de evidência nos dados permite que os gestores questionem as ações dos tribunais de conta.

A limitação também cria dificuldades para que os tribunais façam recomendações às prefeituras. “As recomendações vão ser sempre muito genéricas, muito básicas. Se a gente aponta, mas o apontamento é frágil, a outra ponta vem forte”, explica ele sobre a reação dos gestores, questionando a falta de evidências objetivas das avaliações de controle externo.

Além disso, está em risco a realização de um acordo de cooperação técnica assinado pela Atricon e IRB com o MEC e o Fundo Nacional de Desenvolvimento da Educação (FNDE). O acordo prevê “o **acompanhamento dos planos de educação nacional, estaduais, distrital e municipais** pelos 33 Tribunais de Contas em todo o Brasil e a utilização de instrumentos de monitoramento que concorram para a transparência e efetividade do controle social na utilização dos recursos públicos aplicados em educação”.

Miola enfatiza que o uso dos microdados pelos tribunais traz vantagens para o próprio Inep, uma vez que a fiscalização de perto dos dados no âmbito dos

municípios tem servido ao longo dos anos para ***aprimorar o preenchimento correto dos dados do Censo Escolar*** por parte dos servidores municipais, para evitar possíveis discrepâncias que levem a resultados incongruentes nas bases de dados. Ressalta ainda que o controle externo não se limita a uma análise formal, mas busca verificar se as crianças estão na escola na idade certa e aprendendo, ou seja, ampliando o enfoque puramente quantitativo.

Para o presidente da Atricon e os órgãos de controle externo, as preocupações do Inep em relação à proteção de dados pessoais e o cumprimento da LGPD são válidas e relevantes, e o estudo realizado pela UFMG foi considerado “sólido”. Mas ele faz ressalvas: “É um documento sólido, bem fundamentado e respeitado. Mas faltou a perspectiva do todo. Houve um olhar muito delimitado do ponto de vista do próprio Inep em relação ao assunto. Toda informação como regra é pública. A LAI diz que publicidade é regra e sigilo é exceção. Parece um bordão, mas é esse norte que precisa inspirar o aplicador da lei. Talvez fosse necessário que se desse segurança jurídica aos servidores na tomada de decisão”, diz.

Entre abril de 2023 e janeiro de 2025, a Atricon já realizou pelo menos dez reuniões, encontros ou envio de ofícios a órgãos do governo federal solicitando acesso dos órgãos de controle externo aos microdados. Além do Inep em agosto, outubro e dezembro de 2022,¹⁵ – incluindo um ofício em outubro de 2023,¹⁶ após um ano e meio sem acesso aos dados –, representantes da entidade já se reuniram com a Advocacia-Geral da União (AGU)¹⁷ e até no Palácio do Planalto,¹⁸ para tentar reverter a situação e fazer prevalecer uma leitura menos conservadora da LGPD.

Miola ressalta que os tribunais de conta atualmente já dispõem de processos controlados de acesso a dados sigilosos, incluindo informações bancárias, usados em fiscalizações de controle externo, e que o mesmo também poderia ser aplicado para os microdados educacionais.

A medida, porém, apesar de poder resolver o acesso a alguns órgãos de controle, dificilmente conseguiria estabelecer o mesmo mecanismo para garantir o acesso pelo conjunto de departamentos e áreas de um mesmo tribunal e por outros órgãos de controle ou de auditoria interna das gestões estaduais e municipais, além de ministérios públicos, defensorias e ouvidorias. Não só há inviabilidade administrativa, mas isso

oneraria o Inep de forma desnecessária, uma vez que os dados abertos poderiam estar disponíveis publicamente, atendendo à grande maioria dessa e de outras demandas.

3.3. Poder Legislativo

O Congresso Nacional também reagiu à alteração na política, inclusive com nova legislação, na tentativa de evitar retrocessos na transparência.

» Legislação

Em 8 de março, menos de três semanas após a retirada do ar dos microdados, deputados federais protocolaram o Projeto de Lei (PL) 454/2022,¹⁹ para autorizar o governo federal a “compartilhar e publicizar dados e microdados brutos” do Censo Escolar e do Enem.

Na justificativa do PL, os legisladores explicam que esses dados são essenciais para entender a situação da educação no Brasil, mas sua publicação tem sido adiada ou limitada.

“A restrição impede por completo a compreensão do cenário educacional nacional”, diz o texto, ressaltando que “o atraso no cumprimento das obrigações de proteção de dados pessoais estabelecidas pela LGPD não pode ser utilizado como obstáculo à transparência das informações públicas”.

Para evitar o uso da LGPD como justificativa para limitar a transparência das informações públicas, e levando em conta que o artigo 26 da lei estipula que dados pessoais pseudonimizados podem ser compartilhados quando há previsão legal, os deputados afirmam que o PL foi criado justamente para fornecer respaldo legislativo para o governo federal divulgar os dados do Censo Escolar e do Enem.

O texto original do PL 454/2022 alterava o artigo 5º da Lei de Diretrizes e Bases da Educação (LDB), acrescentando dois novos parágrafos:

§6º O Poder Público é autorizado a compartilhar e publicizar os dados e microdados

brutos coletados no recenseamento escolar de que trata o inc. I do §1º deste dispositivo desde que anonimizados ou pseudonimizados.

§7º O Poder Público é autorizado a compartilhar e publicizar os dados e microdados brutos coletados por meio do Exame Nacional do Ensino Médio, inclusive segmentado por instituição de ensino, desde que anonimizados ou pseudonimizados.

Ele ainda alterava a LGPD em dois artigos. O artigo 14 ganharia dois parágrafos:

§7º O compartilhamento dos dados pessoais coletados no censo escolar, de que trata o §3º do art. 208 da Constituição Federal e o inc. I do §1º do art. 5º da Lei nº 9.394, de 20 de dezembro de 1996, mesmo que de crianças e adolescentes, independe da observância das exigências do §§1º a 6º.

§8º O compartilhamento dos dados pessoais coletados por meio do Exame Nacional do Ensino Médio, mesmo que de crianças e adolescentes, independe da observância das exigências do §§1º a 6º.

Já o artigo 26 também receberia dois novos parágrafos:

§3º O Poder Público é autorizado a compartilhar e publicizar os dados e microdados brutos coletados no recenseamento escolar de que trata o §3º do art. 208 da Constituição Federal e o inc. I do §1º do art. 5º da Lei nº 9.394, de 20 de dezembro de 1996, desde que anonimizados ou pseudonimizados.

§4º O Poder Público é autorizado a compartilhar e publicizar os dados e microdados brutos coletados por meio do Exame Nacional do Ensino Médio, inclusive segmentado por instituição de ensino, desde que anonimizados ou pseudonimizados.

O projeto tramitou nas duas casas legislativas e acabou aprovado – no texto final, o caráter autorizativo da lei se tornou impositivo, ou seja, o governo federal passa a ser obrigado a divulgar os dados, mas ocultando nomes e outras informações que permitam identificar os indivíduos.

Convertido na Lei 15.017/2024,²⁰ o texto foi sancionado pelo presidente Luiz Inácio Lula da Silva em 12 de janeiro de 2025.²¹ A nova lei alterou apenas a LDB em dois artigos.

A nova lei alterou apenas a Lei de Diretrizes e Bases da educação (LDB) em dois artigos.

O artigo 5º ganhou três novos parágrafos:

§ 6º Incumbe ao poder público promover, nos termos de regulamento, o acesso público às informações educacionais do censo anual e dos exames e sistemas de avaliação da educação básica, considerado todo o processo de realização dessas atividades.

§ 7º A organização e a manutenção de sistema de informações e estatísticas educacionais pela União, pelos Estados, pelo Distrito Federal e pelos Municípios, no âmbito da administração direta e indireta, sujeitar-se-ão ao dever de transparência e publicidade como preceitos gerais e ao direito fundamental de acesso à informação de que trata a Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação).

§ 8º Dados e microdados, agregados e desagregados, coletados na execução de políticas educacionais de caráter censitário, avaliativo ou regulatório, serão tratados, divulgados e compartilhados, sempre que possível, de forma anonimizada, observados os parâmetros para anonimização previstos em regulamento. (NR)

O texto também criou o artigo 5º-A, que diz que os três parágrafos acima também se aplicam “às informações educacionais do censo, dos exames e do sistema de avaliação da educação superior”.

Até o início de março de 2025, a legislação estava em fase de regulamentação pelo governo federal.

» Audiências públicas

Em 16 de outubro de 2023, a Comissão de Educação e Cultura do Senado Federal promoveu um debate entre representantes do Inep, da ANPD, da academia e da sociedade civil sobre a divulgação dos microdados educacionais.²² Naquela época, os dados educacionais já estavam fora do ar havia um ano e oito meses.

Participaram do debate:

- » Fábio Pereira Bravin, coordenador-geral de Controle de Qualidade e de Tratamento da Informação (CQTI) do Inep;

- » Lucas Borges de Carvalho, representante da Autoridade Nacional de Proteção de Dados (ANPD);
- » Andressa Pellanda, coordenadora-geral da Campanha Nacional pelo Direito à Educação;
- » Pedro Veloso, coordenador de pesquisa e dados educacionais do Todos Pela Educação;
- » Daniel Funcia De Bonis, diretor de Pesquisa, Dados e Conhecimento da Fundação Lemann;
- » Gabriela Schneider, professora da Universidade Federal do Paraná (UFPR) e representante do Laboratório de Dados Educacionais;
- » Fernanda Campagnucci, diretora-executiva da Open Knowledge Brasil;
- » Pedro Saliba, coordenador de projetos da Associação Data Privacy Brasil;
- » Thiago Alves, professor da Universidade Federal de Goiás (UFG) e representante da Associação Nacional de Pesquisa em Financiamento da Educação (Fineduca);
- » Alexandre Ramos de Azevedo, diretor financeiro da Associação dos Servidores do Inep (Assinep).

Na discussão, Bravin, falando pelo Inep, repetiu as justificativas detalhadas no capítulo anterior, destacando que a preocupação central da autarquia é a possibilidade de que, ao combinar diferentes microdados, seja possível identificar indivíduos, especialmente crianças e adolescentes.

Segundo ele, essa reidentificação pode levar ao uso indevido das informações para fins discriminatórios ou ilícitos, como segregação profissional ou exposição de notas de estudantes, além de poder ser considerada uma violação da LGPD.

O coordenador-geral de CQTI também expressou preocupação de que a reidentificação possa minar a confiança da sociedade no instituto e criar problemas para a cobertura de caráter censitário da coleta dos dados. Em sua opinião, a perda da qualidade da base de dados criaria o risco de inviabilizar a pesquisa no longo prazo, o que seria um problema maior do que restringir o acesso aos dados.

Lucas Borges de Carvalho, representante da ANPD, enfatizou a importância da interpretação da LGPD, e esclareceu que é possível harmonizar a proteção de dados

peçoais com o acesso à informação pública dentro da legislação vigente, uma vez que a lei não se resume apenas a atribuir sigilo a dados pessoais, mas também ampliar a transparência das operações realizadas com esses dados.

Ele lembrou que não há incompatibilidade entre a LGPD e a jurisprudência do STF e citou decisões do Supremo contrárias à imposição de sigilo generalizado sobre todos os processos e favoráveis ao compartilhamento de dados entre órgãos que consolidam o dever de publicidade e transparência, mesmo quando envolve dados pessoais.

“A gente tem que fazer uma interpretação constitucional da LGPD. Ou seja, entender, interpretar a LGPD de forma compatível com a Constituição e também de forma compatível com a jurisprudência do STF. De modo que implementar a LGPD não se resume a atribuir sigilo a dados pessoais. É claro que em alguns contextos e algumas situações o sigilo é uma medida de segurança, uma medida necessária para a proteção desses dados. Mas ele é muito mais do que isso.” (Lucas Borges de Carvalho, ANPD)

Carvalho ressaltou ainda que a LGPD prevê formas de mitigar os riscos de reidentificação de indivíduos, o que demanda uma avaliação de riscos e de impactos no tratamento de dados pessoais, e que a ANPD solicitou ao Inep a elaboração de um Relatório de Impacto à Proteção de Dados Pessoais (RIPD).

Ele também disse que a ANPD divulgará em breve uma nota técnica de avaliação dos RIPDs elaborados pelo Inep e que está disponível para esclarecer dúvidas e contribuir para que haja segurança jurídica entre os servidores federais.

Já os representantes da sociedade civil foram unânimes em defender a revisão da interpretação da LGPD feita pelo Inep e advogar pela retomada da divulgação dos microdados educacionais em prol da transparência, da democratização da educação e da produção de conhecimento essencial para promover a igualdade educacional, avaliar políticas públicas e garantir o controle social.

A falta dos dados limita a capacidade de pesquisadores analisarem a realidade educacional e identificarem desigualdades como as relacionadas à raça e ao desempenho escolar, para colaborarem na proposição de soluções eficazes.

Por sua vez, as Salas Seguras, de acordo com especialistas, não são consideradas uma alternativa adequada, pois impedem análises aprofundadas e têm custos elevados. Uma consequência negativa de impor essa limitação é o risco de mercantilização dos dados, já que o acesso restrito pode favorecer empresas que buscam monetizar essas informações, visto que elas não são mais de acesso público e gratuito.

Em **11 de novembro de 2024**, mais de um ano depois, a **Subcomissão Permanente para tratar do Sistema Nacional de Educação (SNE) do Congresso Nacional** realizou uma audiência pública²³ para debater os sistemas de avaliação. O evento também abordou a falta de publicização dos microdados educacionais. Na ocasião, a nova política do Inep já durava dois anos e nove meses.

Participaram da audiência:

- » Adriana Ventura, deputada federal (Novo/SP) e presidente da subcomissão;
- » Rubens Campos de Lacerda Júnior, diretor de Avaliação da Educação Básica do Inep;
- » José Francisco Soares, professor titular da Universidade Federal de Minas Gerais (UFMG);
- » Joaquim José Soares Neto, presidente da Associação Brasileira de Avaliação Educacional (Abave);
- » Daniel Funcia De Bonis, diretor de Conhecimento, Dados e Pesquisas da Fundação Lemann;
- » Márcio da Costa, diretor de Educação do Instituto Alfa e Beto.

Mais uma vez, os participantes, principalmente representantes da sociedade civil, ressaltaram a necessidade de haver um equilíbrio entre transparência e proteção de dados. A deputada Adriana Ventura destacou que desde o início de 2023 vem cobrando um prazo objetivo para o Inep retomar a divulgação dos microdados educacionais no formato anterior, sem conseguir uma resposta satisfatória.

“Tem havido um movimento quanto aos problemas de falta de dados, da retirada de censo escolar do ar, dos microdados. Eu sei que eles estão avançando nisso, mas faz

anos que estamos esperando e não vejo um horizonte. (...) Somente nesta gestão, nesta Presidência, eu estou cobrando, desde fevereiro de 2023, para ser exata, essa evolução e prazo. O segundo ano já está acabando, e não temos ainda os dados.”
(Adriana Ventura, deputada federal)

Daniel De Bonis, da Fundação Lemann, mencionou a importância de garantir a centralidade da aprendizagem e do direito à aprendizagem, alertando para o risco de retrocessos que diluam o propósito central dos sistemas educacionais em meio a outras variáveis.

Márcio da Costa criticou o acesso restrito aos dados, argumentando que isso prejudica o aprimoramento do sistema como um todo. Daniel De Bonis considerou insuficiente o que o Inep tem feito em relação à disponibilização de microdados, mesmo reconhecendo os esforços em andamento.

Por sua vez, Rubens Lacerda, representando o Inep, mencionou a construção de uma nova plataforma para facilitar a integração entre o Censo Escolar e o Saeb, permitindo aos pesquisadores acessar informações de forma mais facilitada. Ele também informou que o Inep está construindo níveis de adequação ou padrões de desempenho esperados para o 2º, 5º e 9º anos do Ensino Fundamental, com o objetivo de melhorar a comunicação dos resultados das avaliações. E disse que o instituto pretende criar plataformas virtuais para levar essa informação mais próxima das escolas, professores e gestores escolares.

3.4. Poder Judiciário

O Ministério Público Federal (MPF), por sua vez, também se posicionou contrariamente à retirada do ar do formato antigo de microdados.

Em 4 de maio de 2022, o órgão entrou com uma Ação Civil Pública (ACP)²⁴ contra o Inep, afirmando que a justificativa da autarquia para retirada dos microdados não se sustenta porque os dados são anonimizados e a LGPD não proíbe a divulgação de informações de interesse público.

A ação pediu que o Inep voltasse a disponibilizar os microdados “presentes e futuros – de forma a viabilizar que pesquisadores, representantes de instituições privadas, gestores governamentais e a sociedade em geral tenham acesso direto a tais informações e possam realizar diagnósticos, análises, estudos, pesquisas e exercer o controle social em relação à eficácia, eficiência e efetividade de políticas públicas educacionais”.

Foi solicitada uma liminar para que o Inep publique os microdados, sob risco de pagar multa diária. Mas a Justiça indeferiu esse pedido em novembro de 2022.

Segundo o site do Tribunal Regional Federal da 1ª Região (TRF-1 [25]), a ACP seguia em tramitação em fevereiro de 2025, e a última movimentação havia ocorrido em outubro de 2024.

Procurado, o MPF informou, em 19 de fevereiro de 2025, que o processo estava na fase final de andamento. O Ministério Público encaminhou documentos juntados ao processo e o Inep poderia encaminhar a sua manifestação a eles. De acordo com o órgão, isso ainda é possível desde que o juiz encarregado de analisar a ação solicite às partes a apresentação das razões finais, ou que ele dispense essa etapa e parta para o julgamento do processo.

3.5. Poder Executivo

Também o governo federal, por meio de diversos órgãos, se manifestou após a mudança da política.

» Inep

O próprio Inep, devido à repercussão da medida, realizou em 25 de maio de 2022 o *seminário “Controle e tratamento de microdados no Inep”*²⁶ para explicar seu novo posicionamento.

Participaram do evento, entre outros:

- » Eduardo Alencar, assessor de Governança e Gestão Estratégica do Inep;

- » Diego Vasconcelos Costa, gerente de projetos da Autoridade Nacional de Proteção de Dados (ANPD);
- » Mário Sérgio Alvim, Professor da Universidade Federal de Minas Gerais (UFMG), responsável pelo estudo “Privacidade dos Censos Educacionais”;
- » Luiz Felipe Miranda, diretor do Inep;
- » Marcos César, servidor responsável pelo Serviço de Acesso a Dados Protegidos do Inep (Sedap);
- » Guilherme Lambais, pesquisador que utiliza o Sedap desde 2015;
- » Fábio Pereira Bravin, coordenador-geral de Controle de Qualidade e de Tratamento da Informação (CQTI) do Inep.

O professor Mário Sérgio Alvim, da UFMG, explicou a metodologia e os resultados do estudo que baseou a mudança da política do Inep. Ele também usou as *práticas europeias* como um ponto de referência para contrastar com a abordagem do Inep em relação à divulgação dos dados até 2018 e 2019. Segundo o professor, a Europa tende a ser mais rigorosa na proteção de dados e na restrição do acesso a microdados.

Representando a ANPD, Diego Vasconcelos Costa destacou que, na Europa, não há uma regra idêntica à da LGPD, mas sim uma presunção de compatibilidade para o tratamento de dados, baseada em consentimento, interesse público ou legítimo interesse. No Brasil, ele diz que a hipótese de não aplicação da LGPD para fins acadêmicos é uma exceção parcial, vinculada à finalidade acadêmica, e o tratamento de dados pessoais deve estar amparado em uma das bases legais previstas na lei. Ainda reforçou que os fins acadêmicos devem estar diretamente ligados à liberdade de expressão no ambiente acadêmico.

Vasconcelos Costa também baseou sua fala a partir de duas perspectivas principais: o regime estabelecido no artigo 4º da LGPD (que prevê a não aplicação da lei para atividades de tratamento de dados com fins exclusivamente acadêmicos) e as bases legais que permitem estudos por órgãos de pesquisa. O gerente de projetos da ANPD ressaltou a necessidade de estabelecer parâmetros claros para que os dados pessoais identificados sejam usados de forma segura, sem prejudicar a pesquisa científica. Para isso, segundo ele, é necessário

equilibrar o respeito aos direitos fundamentais dos titulares dos dados com a liberdade acadêmica.

“O alcance desse equilíbrio é a observância do princípio da boa-fé e a adoção de salvaguardas técnicas, jurídicas e administrativas. E sempre elas devem ser proporcionais, considerando o risco envolvido naquelas atividades.” (Diego Vasconcelos Costa, ANPD)

Por fim, ele enfatizou a necessidade de distinguir o consentimento exigido pela LGPD daquele previsto em normas éticas de pesquisa, reforçando a importância de alinhar a proteção de dados com as práticas científicas e acadêmicas.

As outras medidas tomadas pelo Inep nos três primeiros anos desde a mudança da política, que foram detalhadas no capítulo 1 deste relatório, envolveram a expansão do Sedap, que hoje conta com cinco núcleos em três estados, e o desenvolvimento de uma plataforma para acesso controlado aos dados, ainda sem previsão de lançamento.

» Autoridade Nacional de Proteção de Dados (ANPD)

O Inep também produziu dois Relatórios de Impacto à Proteção de Dados Pessoais (RIPDs) solicitados pela ANPD após a retirada do ar dos dados.

O pedido da autoridade nacional foi feito em 17 de maio de 2022. Entretanto, os dois relatórios só ficaram prontos em 21 de dezembro (referente aos dados do Enem)²⁷ e 30 de dezembro (dados do Censo Escolar)²⁸ do mesmo ano.

A avaliação da ANPD dos dois relatórios também levou vários meses, sendo concluída apenas em **setembro de 2023**.²⁹

O órgão concluiu que o instituto cumpriu adequadamente as determinações da Coordenação-Geral de Fiscalização (CGF) da ANPD, ao implementar medidas que reduziram os riscos de violação da privacidade e elaborar um RIPD, resultando no encerramento do processo.

Ainda, segundo a ANPD, as medidas possibilitaram ao instituto descrever os processos de tratamento de dados pessoais que poderiam gerar riscos às liberdades civis e aos direitos fundamentais.

“Como se pode observar, a nova abordagem adotada, considerando como medidas (i) a alteração da estrutura dos arquivos e (ii) o acesso controlado por meio do SEDAP, segundo a avaliação do INEP, possibilitou uma redução do risco identificado, levando à aprovação das medidas propostas” (ANPD).

Foi recomendado pelo órgão também a revisão contínua dos RIPDs e a expansão dos canais de acesso controlado, além de sugerir a publicação dos relatórios para demonstrar o compromisso do Inep com a segurança dos dados.

No entanto, a ANPD não adotou, nessa avaliação, uma abordagem de sopesamento de riscos e benefícios, ou seja, não considerou de forma equilibrada os benefícios dos dados abertos em relação aos riscos associados à sua divulgação no formato anterior. Essa ausência de ponderação limita a avaliação, pois impede uma análise mais ampla dos impactos positivos da transparência e do acesso à informação.

» **Controladoria-Geral da União (CGU)**

Poucos dias após a mudança do Inep, a CGU também reagiu, embora sem citar diretamente a autarquia do MEC. Em um enunciado publicado em **10 de março de 2022**, o controlador-geral da União, ministro Wagner de Campos Rosário, reiterou que “A LAI, a Lei nº 14.129/2021 (Lei de Governo Digital) e a Lei nº 13.709/2018 (Lei Geral de Proteção de Dados Pessoais - LGPD) são sistematicamente compatíveis entre si e harmonizam os direitos fundamentais do acesso à informação, da intimidade e da proteção aos dados pessoais, não havendo antinomia entre seus dispositivos”.²⁹

Em **23 de maio de 2022**, a CGU publicou uma nota técnica³⁰ se posicionando contrariamente à interpretação que o Inep fez da LGPD.

O órgão concluiu que a LAI deve ser a principal legislação orientadora, e a não anonimização dos dados não é motivo suficiente para impedir a publicação.

O fato de uma base de dados não estar anonimizada não pode ser usado como único fundamento para que os dados pessoais nela contidos não se tornem públicos, pois a LAI estabelece a publicidade como preceito geral (Art. 3º, I), determina que a divulgação de informações de interesse público deve ocorrer independentemente de solicitações e prevê hipóteses nas quais a publicação de dados pessoais é autorizada. Dessa forma, os gestores deverão realizar avaliação quanto à previsão legal para a publicação de bases de dados e quanto à existência de interesse público e geral preponderante sobre o direito de proteção dos dados pessoais (CGU, Nota Técnica nº 1136/2022/CGAT/DTC/STPC, 2022, p.10).

A controladoria também afirmou que dados já publicados devem continuar a ser atualizados, a menos que haja risco significativo à privacidade, exigindo uma nova avaliação do interesse público.

Referências

1. Disponível em: <https://media.campanha.org.br/acervo/documentos/PautasPolíticas_Inep_Posicionamento_MicrodadosEnem_2022_02_22_FINAL_ok_mcnZoM2.pdf>.
2. Disponível em: <<https://portaliede.org.br/contribuicao/nova-forma-de-divulgar-dados-do-enem-e-do-censo-escolar-pelo-inep-contraria-interesse-publico-e-inviabiliza-diversas-pesquisas/>>.
3. Disponível em: <<https://jeduca.org.br/noticia/nota-sobre-a-retirada-de-microdados-do-site-do-inep>>.
4. Disponível em: <<https://portaliede.org.br/contribuicao/iede-realiza-analise-empirica-sobre-a-supressao-dos-microdados-do-censo-escolar/>>.
5. Disponível em: <<https://ok.org.br/noticia/okbr-e-data-privacy-brasil-lancam-relatorio-sobre-workshop-que-abordou-apagao-de-dados-do-inep/>>.
6. Disponível em: <<https://ok.org.br/wp-content/uploads/2022/09/Relato%CC%81rio-LGPD-e-microdados-OKBR-e-ADPB.pdf>>.
7. Disponível em: <<https://jeduca.org.br/noticia/especialistas-debaterem-alternativas-para-divulgacao-de-microdados>>.
8. Disponível em: <<https://www.youtube.com/watch?v=Clr-bHZKDzI>>.
9. Disponível em: <<https://www.youtube.com/watch?v=klzkXpLAQv4>>.
10. Disponível em: <<https://www.youtube.com/watch?v=b61HeLzasRs>>.
11. Disponível em: <https://associado.anped.org.br/sites/default/files/images/mocao_08_oficio_anped_082-2023_mocao_pela_revisao_da_atual_interpretacao_e_aplicacao_da_lei_geral_de_protecao_de_dados_pessoais_lgpd_pelo_inep.pdf>.
12. Disponível em: <https://fineduca.org.br/wp-content/uploads/2023/12/CARTA-DE-CURITIBA-2023_APROVADA.pdf>.
13. Disponível em: <<https://atricon.org.br/131818-2/>>.
14. Disponível em: <<https://atricon.org.br/wp-content/uploads/2022/03/Nota-Conjunta-no-002-2022-Manifestacao-relativa-a-adequacao-dos-dados-a-LGPD.docx-1-1-1.pdf>>.
15. Disponível em: <<https://atricon.org.br/reuniao-com-inep-trata-sobre-acesso-aos-dados-do-censo-escolar/>>.
16. Disponível em: <<https://atricon.org.br/wp-content/uploads/2023/10/Oficio-no-361-2023-Solicitacao-de-acesso-a-dados-educacionais-do-INEP.pdf>>.
17. Disponível em: <<https://atricon.org.br/dados-educacionais-atricon-trata-sobre-revisao-do-posicionamento-do-inep-na-agu/>>.
18. Disponível em: <<https://atricon.org.br/audiencia-no-palacio-do-planalto-aborda-acesso-aos-microdados-do-censo-escolar/>>.
19. Disponível em: <<https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2317281>>.
20. Disponível em: <<https://www2.camara.leg.br/legin/fed/lei/2024/lei-15017-12-novembro-2024-796541-publicacaooriginal-173508-pl.html>>.
21. Disponível em: <<https://www.camara.leg.br/noticias/1110840-NOVA-LEI-OBRIGA-GOVERNOS-A-DIVULGAREM-DADOS-DE-CENSOS-E-AVALIACOES-EDUCACIONAIS>>.
22. Disponível em: <<https://www.youtube.com/watch?v=330n14vuhq0>>.
23. Disponível em: <<https://escriba.camara.leg.br/escriba-servicosweb/html/74637>>.
24. Disponível em: <<https://www.mpf.mp.br/df/sala-de-imprensa/docs/PRDFMANIFESTACAO147412022.pdf>>.

25. Disponível em: <<https://pje1g.trf1.jus.br/consultapublica/ConsultaPublica/DetalheProcessoConsultaPublica/listView.seam?ca=20f5680a5698a5b4daec1748d338c4e7ae18f5479178b8f1>>.
26. Disponível em: <<https://www.youtube.com/watch?v=S4MRqlligSA>>.
27. Disponível em: <https://download.inep.gov.br/microdados/RIPD_dos_microdados_do_Enem.pdf>.
28. Disponível em: <https://download.inep.gov.br/microdados/RIPD_dos_microdados_dos_Censos_da_Educa%C3%A7%C3%A3o.pdf>.
29. Disponível em: <????>.
30. Disponível em: <https://repositorio.cgu.gov.br/bitstream/1/67735/3/Enunciado_4_2022.pdf>.
31. Disponível em: <https://www.gov.br/cgu/pt-br/assuntos/noticias/2022/05/cgu-monitora-publicacao-de-microdados-de-alunos-pelo-inep/SEI_CGU2376783NotaTcnica.pdf>.

4. IMPACTO

A decisão do Inep de retirar do ar os microdados educacionais para zerar qualquer risco de reidentificação de indivíduos promoveu uma série de impactos nos três últimos anos.

A nova política acendeu o debate sobre a compatibilidade da LGPD e da LAI – ou seja, sobre o equilíbrio entre transparência e privacidade na execução de políticas públicas. Ela também teve efeitos qualitativos diretos na produção científica na área educacional (e em outras áreas), no uso da LAI para obter dados e na busca pelo Sedap.

Houve ainda impacto indireto e subjetivo, da ordem qualitativa, como veremos a seguir.

4.1. Publicação de artigos e pesquisas acadêmicas

Não há uma maneira simples de verificar exatamente quantos pesquisadores usavam os microdados públicos retirados do site do Inep no formato vigente entre 2009 e 2022. Seria necessário um esforço de leitura e tabulação das metodologias de milhares de trabalhos acadêmicos para entender a fonte exata dos dados.

Mas um levantamento feito pela OKBR no início de 2025, a partir de resultados de diferentes buscas realizadas no sistema Google Acadêmico,¹ nos oferece algumas pistas.

A pesquisa foi feita de forma manual, comparando a mesma seleção de buscas textuais, sempre de publicações em português, em recortes anuais dos últimos 15 anos entre 2010 e 2024.

Os termos usados, sempre entre aspas para delimitar os resultados e evitar falsos positivos, foram:

Grupo 1 de termos

- » "dados do censo escolar"
- » "dados do censo da educação"
- » "microdados do inep"
- » "dados do inep"
- » "dados do saeb"
- » "dados do enem"
- » "dados do exame nacional do ensino médio"

Para efeito de comparação, buscou-se também por artigos que citavam os termos:

Grupo 2 de termos

- » "Acesso a Dados Protegidos" Inep
- » Sedap Inep

A partir da quantidade de resultados informada na busca de cada termo em cada ano, uma tabela foi criada para sistematizar o número de publicações.

Vale ressaltar que duas buscas usando termos parecidos, como "dados do enem" e "dados do exame nacional do ensino médio" e, de forma semelhante, a busca por "*Acesso a Dados Protegidos*" Inep e *Sedap Inep*, podem retornar a mesma publicação de forma repetida. Por isso, o total de menções não necessariamente representa o número individual de artigos.

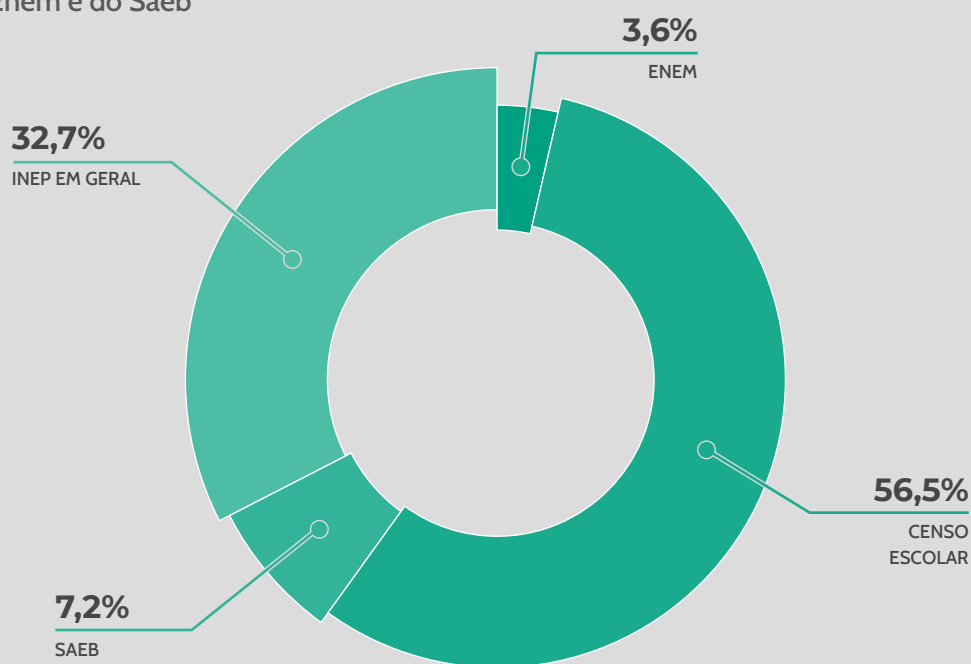
O método de busca aplicado foi sempre o mesmo para todas as consultas, em todos os anos. Supondo que não houve nenhum grande evento capaz de alterar significativa e amplamente a forma de escrita de pesquisadores e pesquisadoras, a margem de inclusão de menções repetidas se aplica de forma relativamente homogênea no volume de artigos, em todos os anos. Por isso, podemos considerar que a análise está sendo feita com um universo total de publicações que atendem aos critérios da coleta e cujas repetições de termos não tenham relevância significativa para a análise do volume de publicações entre diferentes anos, como a explorada aqui.

Em uma primeira soma da tabela criada, verificamos que, para o **Grupo 1** de termos, as buscas resultaram em **28.667 menções entre 2010 e 2024**, uma média de 1.911 por ano.

Dessas, a base de dados mais citada foi a do Censo da Educação Básica, ou Censo Escolar, que respondeu por 56,5% dos resultados. A busca por publicações citando dados ou microdados do Inep concentrou 32,7% dos resultados, seguida pelos resultados citando o Sistema de Avaliação da Educação Básica (Saeb), com 7,2% e o Enem, com 3,6%.

GRÁFICO 1 – MENÇÕES EM ARTIGOS, EM PORTUGUÊS, PUBLICADOS ENTRE 2010 E 2024

Resultados de buscas por termos relacionados aos dados do Censo Escolar, do Enem e do Saeb



Fonte: Levantamento Open Knowledge Brasil feito a partir de dados do Google Acadêmico em fev/2025.
Nota: Os dados referentes a 2024 podem sofrer alterações devido a possíveis atrasos de notificação.

Já a distribuição dos resultados por ano de publicação de menções em artigos e trabalhos científicos mostra uma evolução compatível com o avanço das políticas de transparência do Inep, das técnicas de ciência de dados, da expansão de usuários das bases de dados e da abrangência do Google Acadêmico.

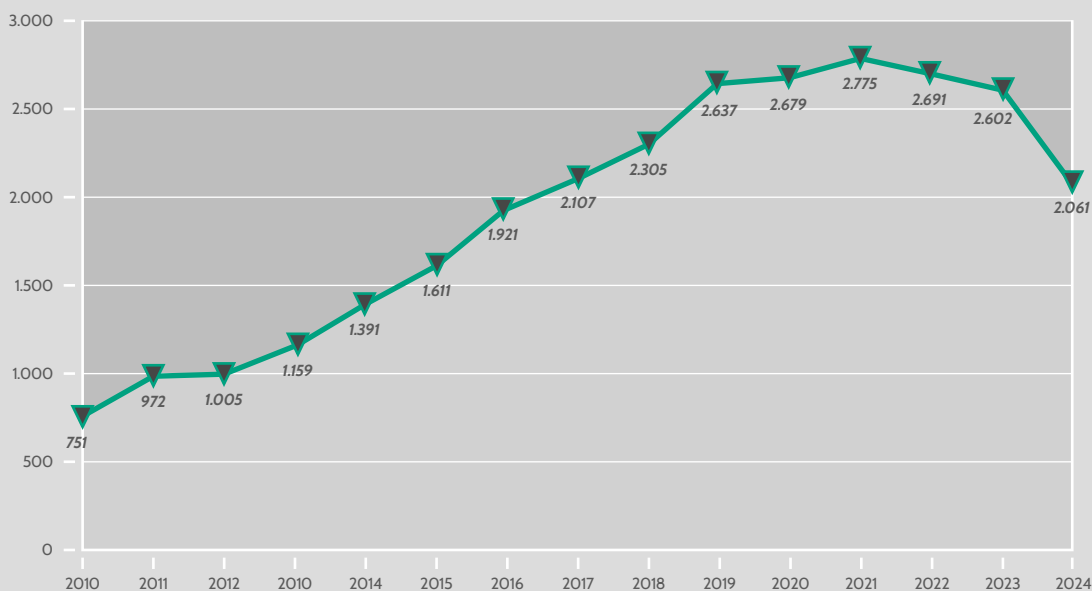
Entre 2010 e 2021, a quantidade de resultados das buscas do Grupo 1 cresceram sistematicamente, saltando de 751 para 2.775, uma alta de 269,5% em 12 anos. A curva, porém, tem caído de forma sistemática desde 2022, o ano da mudança na política de transparência do Inep.

Em 2022 e 2023, o total de resultados caiu 3% ao ano, chegando a 2.602, um patamar menor que o de 2020. Já de 2023 para 2024, a queda foi mais ampla, de 21%, embora essa comparação ainda deva ser considerada preliminar, devido a possíveis atrasos entre a publicação dos artigos e seu compartilhamento em sites que integram o universo de indexação do Google Acadêmico – que não divulga, em sua metodologia e orientações às pessoas pesquisadoras, a média de tempo entre a publicação e a incorporação do artigo no sistema.²

Mesmo assim, até meados de fevereiro de 2025, seguindo a mesma metodologia, a busca resultou em 2.061 publicações feitas em 2024 até agora, sendo o valor mais baixo em oito anos.

GRÁFICO 2 - ARTIGOS EM PORTUGUÊS POR ANO DE PUBLICAÇÃO

Resultados de buscas por termos relacionados aos dados do Censo Escolar, do Enem e do Saeb.



Fonte: Levantamento Open Knowledge Brasil feito a partir de dados do Google Acadêmico em fev/2025.

Nota: Os dados referentes a 2024 podem sofrer alterações devido a possíveis atrasos de notificação.

TABELA 1 - MENÇÕES EM PUBLICAÇÕES INDEXADAS PELO GOOGLE ACADÊMICO

ano	"dados do censo escolar"	"dados do censo da educação"	"microdados do inep"	"dados do inep"	"dados do saeb"	"dados do enem"	"dados do exame nacional do ensino médio"	total
2010	219	101	3	290	115	22	1	751
2011	305	138	5	391	111	17	5	972
2012	325	152	6	377	114	27	4	1005
2013	379	196	16	436	108	20	4	1159
2014	491	237	18	480	121	40	4	1391
2015	566	290	20	551	121	44	19	1611
2016	718	328	24	647	137	56	11	1921
2017	838	383	28	607	157	79	15	2107
2018	938	439	25	696	108	81	18	2305
2019	1010	520	38	807	156	85	21	2637
2020	1010	573	40	798	165	80	13	2679
2021	1050	603	45	816	170	83	8	2775
2022	1040	585	42	751	166	90	17	2691
2023	1060	537	57	683	175	72	18	2602
2024	864	418	30	536	135	60	18	2061

Já para o Grupo 2 de termos, buscando especificamente trabalhos acadêmicos que citam diretamente o Serviço de Acesso a Dados Protegidos do Inep (Sedap), uma busca pelos dois termos diferentes no mesmo período de 15 anos resultou em apenas 145 publicações (que, como no Grupo 1, podem aparecer mais de uma vez nesse total).

É importante notar que o Sedap foi criado em 2014 pelo Inep, o que já torna natural a inexistência de menções a ele em anos anteriores.

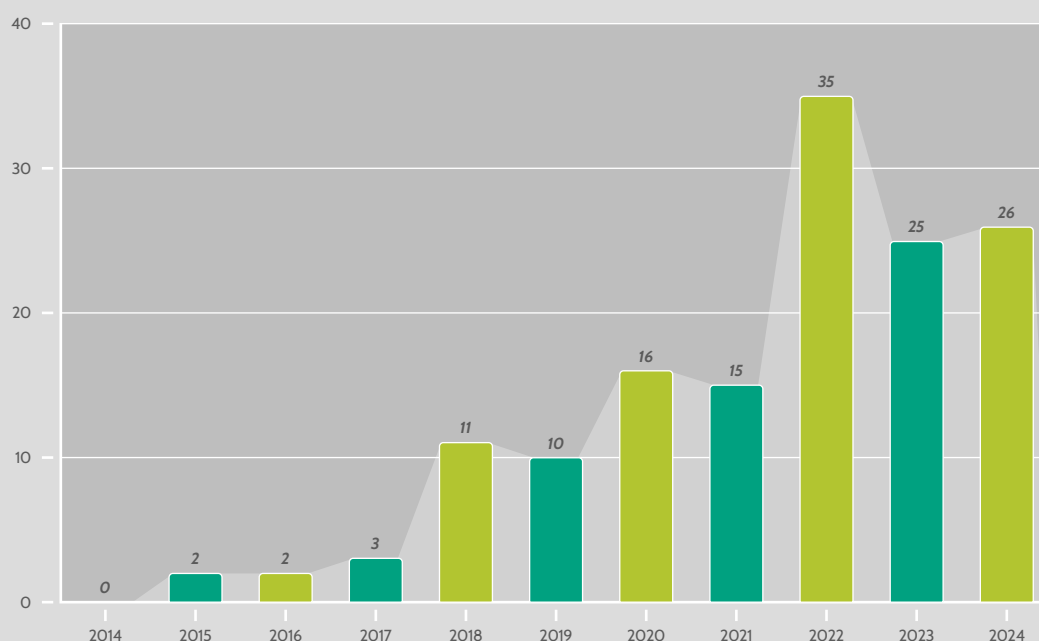
De todas as formas, o resultado das buscas é bem menor do que no levantamento feito com termos do Grupo 1 (as 145 publicações resultantes representam apenas 0,5% dos resultados obtidos nas buscas do Grupo 1).

Além disso, a busca de termos do Grupo 2 não delimita o teor ou área de conhecimento na qual cada pesquisa está inserida. Portanto, os resultados podem incluir publicações produzidas após uma visita das pessoas autoras ao Sedap para coletar dados educacionais ou mesmo serem de outras áreas, como Transparência ou Proteção de Dados Pessoais, que citam o Sedap como exemplo de acesso controlado a informações protegidas – um fenômeno que também se aplica.

Mas a distribuição dos resultados pelo ano de publicação da pesquisa aponta para uma tendência de alta em relação à série histórica justamente em 2022, ano que coincide com a nova política de proteção de dados do Inep. Entre 2021 e 2022, a quantidade mais que dobrou, de 15 para 35.

GRÁFICO 3 – ARTIGOS EM PORTUGUÊS POR ANO DE PUBLICAÇÃO

Resultados de buscas por termos relacionados ao Sedap/Inep.



Fonte: Levantamento Open Knowledge Brasil feito a partir de dados do Google Acadêmico em fev/2025.

Nota: Os dados referentes a 2024 podem sofrer alterações devido a possíveis atrasos de notificação.

Além disso, mesmo com a possibilidade de atraso de notificação, os resultados para 2024 já superaram os de publicações feitas em 2023. A média de publicações por ano aumentou de 8,4 entre 2015 e 2021 para 28,6 entre 2022 e 2024 (alta de 240%).

Os números acima, apesar de se limitarem apenas à produção acadêmica em português presente no índice do Google Acadêmico, mostram quantitativamente o impacto de não ter as bases no maior nível de desagregação disponíveis para acesso público e gratuito.

Já o Sedap, segundo dados produzidos e divulgados pelo próprio Inep, viu a quantidade de solicitações de acesso à Sala Segura mais que dobrar entre 2023 e 2024. Porém, esse número não chega perto da totalidade de publicações no Google Acadêmico: em 2022 e 2023, foram 38 e 37 processos, respectivamente, incluindo os que tiveram o pedido indeferido. Isso evidencia o quanto do acesso à informação vem ocorrendo, na prática, e com impactos substantivos cada vez maiores para a pesquisa em Educação no Brasil. No ano de 2024, o total de solicitações de acesso à Sala Segura ao Sedap saltou para 76, trazendo à tona a questão adicional: a insustentabilidade da Sala Segura em termos de escala.

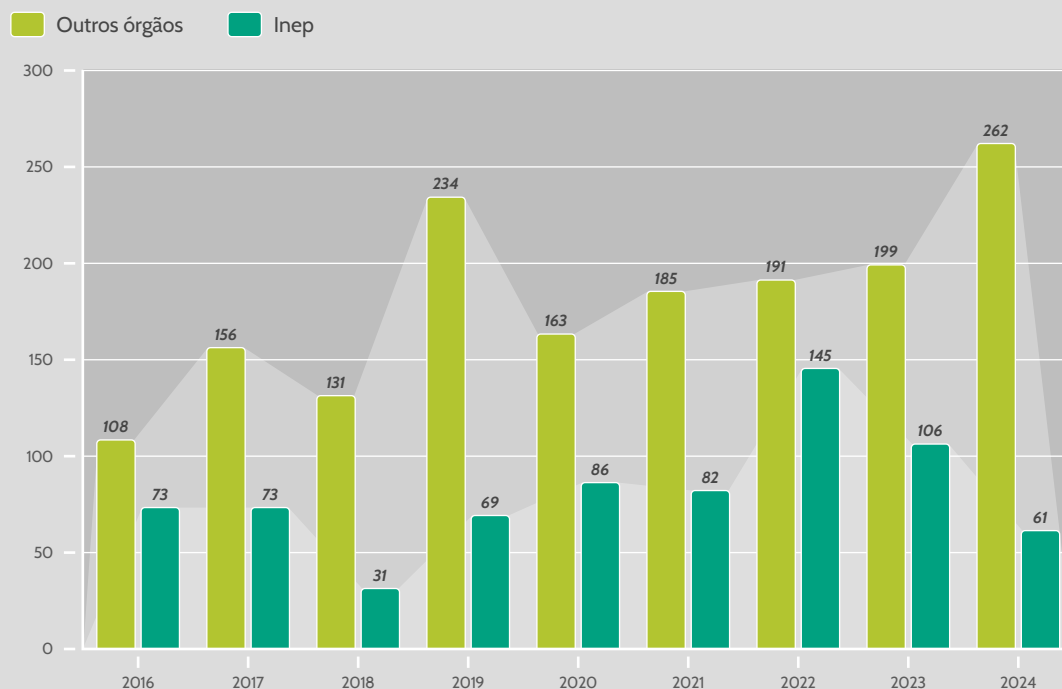
4.2. Pedidos via Lei de Acesso à Informação

Outro levantamento quantitativo feito pela Open Knowledge Brasil para o presente relatório usa como fonte o sistema da Controladoria-Geral da União (CGU) de acesso a pedidos e respostas via LAI. Essa lei foi aprovada em 2011, mas o sistema fornece detalhes sobre todos os pedidos feitos a partir de 2015.

Buscando na base completa de pedidos por solicitações que foram enviadas especificamente ao Inep, e que contêm, na demanda, a palavra “microdados”, os resultados também mostram como a tendência de alta nos pedidos coincide com a nova política.

GRÁFICO 4 – PEDIDOS VIA LEI DE ACESSO À INFORMAÇÃO

Solicitações contendo as palavras 'microdados' ou 'dados'.



Fonte: Levantamento Open Knowledge Brasil com dados abertos da CGU (jun/25).

Entre os principais achados, a análise de 765 pedidos recebidos pelo Inep, contendo a palavra “microdados” e disponíveis no sistema da CGU em junho de 2025, verificou os seguintes impactos:

→ **60,7% de aumento de pedidos via LAI**

Entre janeiro de 2015 e 18 de fevereiro de 2022 (data em que os microdados foram retirados do ar), a **média de pedidos por mês** que se encaixam nos critérios acima foi de 6,1. Entre 18 de fevereiro de 2022 e dezembro de 2024, esse número **saltou** para 9,7 pedidos por mês, representando um crescimento de **60,7%**.

Em 2022, ano da mudança na política, o Inep chegou a responder sozinho por 43% de todos os pedidos enviados via LAI por cidadãos solicitando “microdados”. Foram 145 pedidos diretamente ao Inep, e 191 para a soma dos demais 125 órgãos integrantes do sistema da CGU.

→ *53,2% de aumento na prorrogação de prazos*

Vale lembrar que cada pedido é lido e respondido de forma individual por servidores do próprio Inep. Com isso, o aumento da demanda também impactou a eficácia do Serviço de Informação ao Cidadão (SIC) do Inep.

Considerando os pedidos contendo a palavra “microdados” feitos ao Inep entre janeiro de 2015 e 17 de fevereiro de 2022, **17,7%** tiveram o prazo inicial de resposta (de 20 dias corridos) prorrogado.

Já considerando os pedidos feitos entre 18 de fevereiro de 2022 e dezembro de 2024, mais de um quarto dos cidadãos (**27,2%**) precisaram esperar além do prazo inicial para receber alguma resposta. Podemos inferir que houve uma alta de **53,7%** em comparação com o período anterior.

→ *Aumento de informações negadas por ‘trabalho adicional’*

A LAI de Acesso à Informação tem algumas formas de classificar as respostas. No geral, o pedido pode ter acesso à informação **“concedido”, “parcialmente concedido” ou “negado”**.

Seguindo esses conceitos, o universo considerado no levantamento manteve as médias de atendimento do Inep antes e depois da mudança da política: de cada 100 pedidos, entre 80 e 84 têm “acesso concedido”, entre 9 e 10 têm “acesso parcialmente concedido”, 6 têm “acesso negado” e o restante é classificados como “informação inexistente”, “pedido duplicado”, “não se trata de solicitação de informação” ou pedido sobre um assunto pelo qual o órgão não tem competência para responder.

Por outro lado, embora representem uma pequena minoria do total, desde 18 de fevereiro de 2022 houve aumento de **50%** na proporção de pedidos que, na posição do Inep, “exigem tratamento adicional de dados”, “demandam mais tempo para produção” ou são “desproporcionais ou desarrazoados”.

Também em comparação com junho de 2025, aumentou (nesse caso, em **35,5%**) a proporção de demandas que foram classificadas pela autarquia como tendo solicitado “dados pessoais”.

4.3. Uso dos microdados em geral

Há ainda impactos mais subjetivos da retirada do ar dos microdados. Esses só podem ser medidos por uma abordagem qualitativa, mas revelam prejuízos substanciais.

Um exemplo é o enfraquecimento das pesquisas educacionais em geral. Assim como vimos, no capítulo 1, a professora **Gabriela Schneider**, da UFPR, que passou a dedicar sua carreira na pós-graduação à análise de dados educacionais por causa do incentivo do governo federal em oferecer bolsas – no âmbito do programa Observatório da Educação –, a ausência dos microdados também provoca o efeito oposto.

Ernesto Martins de Faria, fundador do Iede e um dos principais pesquisadores na área, começou sua formação em Economia e Econometria, mas acabou encaminhando sua pesquisa no campo do impacto socioeconômico no desempenho dos estudantes. Segundo ele, a existência dos microdados foi crucial para isso – entre seus estudos está o uso dos dados para criar níveis socioeconômicos e poder avaliar a qualidade da educação ao agrupar estudantes e redes nessa escala. Hoje, o risco é de a pesquisa na área estagnar, ou até retroceder, dados os novos obstáculos de acesso à informação necessária para os estudos.

“Pensando no Brasil, na pessoa fazendo iniciação científica, monografia, dissertação, isso acaba direcionando para que os professores trabalhem menos com educação, porque eles não têm base para os alunos poderem trabalhar. Ele vai começar a olhar para outras áreas, vai olhar para o mercado de trabalho, para a saúde. Isso acaba retirando o incentivo de fazer pesquisa em educação.” (Ernesto Martins de Faria, Iede)

Observa-se que, além de afastar pesquisadores, o próprio Inep acaba se afastando de um posto que assumiu como modelo, entre órgãos públicos, de transparência, organização e disseminação de microdados.

Nos anos em que ampliou a divulgação das bases de dados, o Inep também modernizou de forma considerável a forma como eles eram apresentados à sociedade, com arquivos muito bem trabalhados e esmero na produção de metadados (a documentação que acompanha os microdados e contém detalhes e orientações para os cidadãos usarem o material de forma adequada).

Considerados modelos para outros órgãos, os microdados do Inep também eram usados na democratização e capacitação de usuários iniciantes na área de Ciências de Dados. Alguns exemplos são cursos de formação que optaram por ensinar técnicas usando como base os dados do Inep. O curso “SQL para Jornalistas” da Associação Brasileira de Jornalismo Investigativo (Abraji)³ teve pelo menos quatro edições entre 2016 e 2017, capacitando mais de 230 jornalistas antes de se tornar fixo na plataforma online de formação da associação. As edições do curso usavam, entre outras bases de dados, os censos da Educação Básica e da Educação Superior, adequados ao nível iniciante de aprendizado.

Já a Escola de Dados da Open Knowledge Brasil ofereceu dois cursos: um de “Jornalismo de dados para coberturas locais”, que dedicou um módulo específico para explicações sobre como usar as bases de dados do Inep,⁴ e um chamado de “Análise de Dados Educacionais: aplicando evidências na gestão pública”, com até 320 vagas e voltado para gestores públicos da área de Educação.⁵

Para os usuários das bases que se mantiveram na pesquisa de dados educacionais, a limitação de dados abertos também desfavorece a reprodutibilidade científica. Uma vez que os dados estão disponíveis publicamente, outros pesquisadores podem replicar estudos, verificar resultados e propor novos métodos de análise, fortalecendo a robustez das pesquisas acadêmicas.

Gabriela Schneider conta que, com o trabalho da plataforma do custo aluno-qualidade mantida pela UFPR, os pesquisadores já foram capazes inclusive de detectar erros nas bases de dados do Inep, e alertar o órgão para que ele corrigisse as informações e aprimorasse a qualidade dos dados.

4.4. Atuação da sociedade civil

Além da produção científica nas universidades e da produção de reportagens sobre políticas públicas educacionais, outro grande uso dos microdados é feito pela sociedade civil organizada.

Associações, organizações e institutos atuantes no campo da educação mantêm departamentos inteiros dedicados ao acesso e análise de dados para produção de conhecimento e contribuição ao debate sobre a qualidade do ensino.

Uma dessas entidades é o Instituto Ayrton Senna (IAS). A pesquisadora Gabriela Cáceres ocupa o cargo de coordenadora de Monitoramento e Avaliação do IAS e explica que a alteração na política de divulgação desses dados impactou negativamente seu trabalho.

Cáceres diz que utiliza os microdados para criar diagnósticos e entender o cenário da educação pública brasileira, analisando como os resultados se distribuem no país. Ela destaca ainda que a capacidade de desagregar dados permite identificar desigualdades e disparidades em diferentes níveis.

“Antes a gente tinha os dados de fato microdados, no nível mais baixo [desagregado], das pessoas que estavam na escola, fossem elas estudantes, professores, diretores. A gente sabia quem eram, seu gênero, cor ou raça, idade, a gente podia entender essas características. Hoje, com essas mudanças, a gente não consegue mais.”
(Gabriela Cáceres, IAS)

Segundo ela, os dados agora são agregados por escola e etapa de ensino, e não por ano escolar. Com isso, não é mais possível identificar características individuais para analisar, por exemplo, a distorção idade-série por grupo. “Faz com que a gente esteja restrito na realidade em termos de diagnóstico e análise. Será que meninas negras têm distorção maior que meninas ou meninos brancos? Não consigo mais saber”, explica.

Apesar de o Inep seguir chamando as bases de dados no formato criado em 2022 de “microdados”, Cáceres diz que, conceitualmente, **a definição não é precisa**. Isso

porque os microdados representam os dados em seu estágio mais desagregado, o que, no caso do Censo Escolar, seria cada aluno ou cada docente.

No entanto, desde 2022 o Inep divulga apenas dados de cada escola, criando colunas novas nessa tabela com algumas somas, como o total de estudantes do Ensino Fundamental I, o total de alunos do gênero feminino ou o total de docentes da escola. Ou seja, essas são agregações, justamente o oposto de dados desagregados.

A coordenadora de Monitoramento e Avaliação do IAS ressalta ainda que a sinopse estatística (os arquivos em formato XLSX onde o Inep inclui diversas tabelas com dados agregados por município) hoje em dia apresenta dados mais desagregados que os chamados “microdados” do Inep, o que não deveria acontecer. Nesse caso, as sinopses indicam, por exemplo, o número de estudantes por idade em cada ano escolar, algo que não é possível com os microdados atuais, que fornecem dados apenas por faixa de idade.

O impacto principal da mudança da política, para a pesquisadora do IAS, é o aumento da dificuldade para o Brasil combater desigualdades educacionais, uma vez que não é possível saber se os estudantes de baixo nível socioeconômico ainda sofrem com baixa qualidade de ensino.

Cáceres diz que entende a preocupação do governo em proteger os dados de pessoas com menos de 18 anos", mas, na sua opinião, existem outras formas de mascarar os dados para não afetar as pesquisas. Também acredita que a postura do Inep na criação de Salas Seguras em mais partes do país tem sido passivo demais, dado o tamanho do impacto da retirada dos dados para a produção científica. Para ela, o órgão deveria estar ativamente buscando parcerias com universidades e centros de pesquisa para o acesso aos dados, em vez de esperar que estes o procurem.

4.5. Execução de políticas públicas

Tanto órgãos do governo federal quanto das esferas estadual e municipal também figuravam no rol de usuários dos microdados abertos do Inep. Com a

mudança na política, vários deles tiveram que seguir pela alternativa e tentar o acesso ao Sedap.

Em outubro de 2023, a Secretaria de Educação Continuada, Alfabetização de Jovens e Adultos, Diversidade e Inclusão (Secadi), vinculada ao MEC, solicitou e acessou a Sala Segura para uma pesquisa sobre “Equidade Educacional na Educação Básica e Superior”.

Outros ministérios também dependem dos dados do Inep. Pelo menos dois exemplos foram confirmados pela Open Knowledge Brasil na produção desse relatório:

Segundo relatórios do próprio Sedap, a Secretaria de Avaliação, Gestão da Informação e Cadastro Único (SAGICAD), do Ministério do Desenvolvimento e Assistência Social, Família e Combate à Fome (MDS), solicitou em dezembro de 2023 acesso à Sala Segura para obter dados para a pesquisa “Oferta e Demanda por Pré-Escola: uma análise a partir do Cadastro Único”.

Por outros canais de atendimento, o acesso aos dados ficou comprometido em pelo menos uma ocasião: via Lei de Acesso à Informação, o Departamento de Avaliação, Monitoramento, Estudos e Informações Estratégicas (DAMEI), do Ministério do Desenvolvimento Agrário e Agricultura Familiar (MDA), informou que a pasta solicitou por e-mail ao Inep, em outubro de 2024, o “acesso a dados com o objetivo de apoiar no dimensionamento do público de assentados da reforma agrária que se encontram em cursos de nível superior”. Ainda segundo o MDA, “essas informações seriam destinadas a apoiar a reflexão e possível elaboração de políticas educacionais específicas para esse público-alvo”. Porém, o órgão disse que “não teve um retorno direto por parte do Inep” e, “desse modo, [o pedido] não resultou na produção de estudo, análise ou documento formal específico do MDA”.

Já na esfera estadual, em abril de 2024 um servidor da Secretaria Estadual de Educação de Minas Gerais (SEE/MG) esteve no Sedap para coletar dados para uma avaliação sobre “equidade e educacional a partir da vulnerabilidade da trajetória escolar de 2008 a 2022 para a rede estadual de Minas Gerais”.

4.6. Limitações do Sedap enquanto alternativa

Mesmo com mais salas em três estados brasileiros, Gabriela Cáceres explica que o Sedap não é um substituto à altura da divulgação pública dos microdados.

A primeira limitação é justamente a falta de microdados abertos para permitirem uma primeira análise exploratória dos dados, etapa fundamental para que a pessoa pesquisadora consiga estruturar um projeto de pesquisa nos moldes do que o Sedap exige para deferir a solicitação de acesso.

As salas de sigilo são *burocráticas e restritivas por natureza*. Além das limitações financeiras e de tempo para dedicar a visitas presenciais para a análise, outras questões comprometem a própria qualidade da pesquisa, segundo relatos da própria pesquisadora do IAS, que já fez uso do serviço.

“O principal ponto é que você precisa fazer todas as análises lá dentro”, explica **Gabriela Cáceres**. “Eu acho que isso é mais complexo, porque muitas das vezes em que você está analisando um dado, faz parte do processo você discutir, você maturar aquela análise. E [na Sala Segura] você não consegue ter esse tipo de discussão.”

Segundo ela, os pesquisadores têm um determinado período de tempo para fazer todas as análises. Não é permitido entrar com nada na sala, nem papel com o código de programação necessário para analisar as bases ou para fazer anotações. “E você tem que usar o computador deles, então você tem que levar um pen drive, você não pode nem usar o seu próprio computador. É complicado.”

O economista **Guilherme Lambais**, professor assistente da Universidade Lusíada de Lisboa, já usou pelo menos duas vezes a Sala Segura do Inep, em Brasília. Por conta disso, ele foi convidado pelo Inep para dar seu depoimento sobre o Sedap no seminário “Controle e tratamento de microdados no Inep”, em 25 de maio de 2022.⁶

O pesquisador afirmou na época que, ao longo dos anos, o Inep aprimorou os recursos e a segurança da sala.

"Foi notável a melhora dos processos da Sala Segura. Antigamente, (...) era mais seguro fisicamente, tinha que passar pelo scanner corporal. Uma vez ele me pegou com papel de bala no bolso, não podia nem entrar com isso. Então apesar de você poder ver outras informações você não tinha como carregar nenhuma informação. Hoje a anonimização é bem mais segura, não precisa mais de um scanner corporal, facilitou a fazer pesquisa ali." (Guilherme Lambais, pesquisador)

Ele também informou que fez sua graduação em outra área de atuação (na Economia) e usando dados de outro órgão, que ele não identificou. E, por causa da falta de divulgação dos microdados, tanto ele quanto outros pesquisadores acabaram desistindo de tentar acessar informações daquela área e migraram para outros campos de pesquisa.

Guilherme contou ainda que, em diversas ocasiões, seu uso da Sala Segura foi como colaborador, auxiliando outros pesquisadores que estavam fisicamente longe de Brasília. Ele citou dois casos desses, nos quais atuou como intermediário para acessar a Sala Segura e garantir que os pesquisadores conseguissem acesso aos dados sem precisar se deslocar até a sede do Inep.

Um dos casos envolvia uma doutoranda brasileira que estudava na Universidade de Harvard e, no outro, um projeto de pesquisa vinculado ao Centro Brasileiro de Análise e Planejamento (Cebap) e às universidades de Harvard e de São Paulo (USP).

Gabriela Schneider, professora da UFPR, também fez uso do Sedap em Brasília. Mas, após uma experiência recente, ela explica que passou a desenhar estratégias de pedir acesso à sala apenas para análises pequenas e pontuais. "Você não pode ir com um objetivo muito grande, eu fui pegar uns dados muito específicos. Consegui coletar", relata.

Mesmo assim, esse acesso levou quatro dias úteis, mas exigiu cinco diárias de hotel. "Fiquei cinco dias lá. Tive que ir na quinta e voltar na terça, para não perder aula [que daria] na quarta. Fui na quinta de manhã, fiquei [na Sala Segura] quinta e sexta, passei o fim de semana em Brasília para usar [a Sala Segura] na segunda e na terça e voltar [para Curitiba] na terça à noite."

Ela explica que a sala pode ser usada por até duas pessoas concomitantemente, em pesquisas e máquinas distintas, mas “o computador não é dos melhores”, diz. “Na sexta mandei rodar uma análise e o computador travou. Fui embora porque fecha mais cedo, quando voltei na segunda ele ainda estava travado. Um servidor ligou para a equipe de TI, passaram duas horas e só então descobri que eu poderia reiniciar o computador eu mesma”, relembra a professora.

Guilherme Lambais, da Universidade Lusíada de Lisboa, também sugere que o Inep aprimore o poder computacional do servidor do Sedap, “para lidar com as bases que são muito grandes”. Segundo ele, o computador na sala “às vezes dá uma engasgada para rodar os dados”.

Ele também recomenda que o Inep mantenha dentro da sala um manual para que os pesquisadores possam aplicar os comandos da linguagem e programação SAS, a única aceita pelos computadores para o processo de baixar os dados do servidor. “Não é muita gente que conhece o programa, eu mesmo não conhecia. A gente teve que aprender na marra na primeira vez.”

Após o acesso aos dados, o pesquisador diz que os computadores oferecem os programas Stata e RStudio para que os usuários realizem as análises necessárias.

Para efeito de comparação, o Ipea diz que sua sala sigilosa dispõe de “computadores de alta capacidade com *softwares* estatísticos” variados, permitindo que os pesquisadores possam fazer consultas usando a linguagem de programação de sua escolha. O espaço também permite que essas máquinas sejam acessadas remotamente, ou seja, usando o computador da própria pessoa pesquisadora, mediante prévia solicitação.

No entanto, pelas regras do Sedap, os pesquisadores não saem com as análises finais no mesmo dia. Primeiramente, uma análise é feita para garantir que nenhum dado pessoal será transferido para fora da Sala Segura. Esse processo pode demorar dias ou semanas para ser finalizado.

Para a professora **Gabriela Schneider**, da UFPR, esse é mais um motivo pelo qual o serviço deve ter como finalidade o atendimento a levantamentos de dados muito específicos e que exigem de fato o acesso a dados pessoais. “O Inep tem feito iniciativa de criar Salas Seguras em outros lugares, mas eu tenho dito isso em vários eventos: ***pesquisa não se faz em Sala Segura***. Às vezes você precisa fazer teste, refinar”, ressalta.

Por causa das novas limitações, a professora diz que tem desestimulado seus orientandos a fazer pesquisas quantitativas. O laboratório do qual ela é coordenadora no Paraná, por exemplo, deixou de poder atualizar plataformas de acesso aberto para a sociedade e gestores municipais e estaduais.

Um deles, o SimCAQ (Simulador do Custo Aluno-Qualidade), é desenvolvido pela UFPR em parceria com a Universidade Federal de Goiás (UFG). Seu surgimento se deu a partir de pesquisas acadêmicas realizadas em 2010 e 2011 a respeito do financiamento educacional, em uma tentativa de entender quanto é preciso investir por aluno para garantir uma educação pública com um patamar mínimo de qualidade.

Parte da pesquisa já havia sido financiada pelo programa Observatório da Educação. ***Desde 2015 o Ministério da Educação (MEC) financia a plataforma***, que permite a qualquer pessoa simular quanto deve ser o investimento em cada rede de ensino, a partir de dados do Censo Escolar (total de matrículas e escolas) e do valor do Fundeb calculado para cada estudante, de acordo com a etapa de ensino em que ele está matriculado, e se a escola fica na zona urbana ou rural.

Hoje, o projeto é feito com o ***trabalho de mais de 20 pessoas*** das duas universidades.

“A gente criou uma rotina com dados desde 2007 e atualizados todo ano. [Com a retirada do ar dos microdados do Inep] a gente passou quase um ano e meio tentando conseguir os dados. O SimCAQ tinha informações como o número de matrículas por idade, a gente não consegue mais fazer isso. Ele calculava o tempo integral de forma diferente [do Inep e do MEC], a gente calculava pelas turmas. [A decisão] afeta totalmente, porque você para de ter [esses dados].” (Gabriela Schneider, UFPR)

Em 2024, após a mudança na política do Inep, a plataforma financiada pelo MEC e executada pelas duas instituições federais precisou passar por um **amplo processo de reformulação**, e a nova versão lançada no ano passado apresenta “limitações para o diagnóstico das condições da oferta de ensino devido à ausência de dados”, segundo a página.

“Entendo que tem algumas coisas que você não pode acessar, como o CPF [dos indivíduos]. Mas é diferente do que você tinha nos microdados. Você poderia tirar algumas informações e deixar o restante. Mas eles tiraram tudo”, conta a professora da UFPR.

Para ela, ter que acessar os dados via Sedap a cada nova edição do Censo Escolar inviabilizaria o projeto, criado justamente para a elaboração de produtos e conhecimento a partir de dados que eram públicos e de interesse público.

Porém, Schneider alerta que muitas outras pesquisas acabam inviabilizadas se dependerem da Sala Segura. Ela cita como exemplo o caso de uma orientanda sua de mestrado que defendeu a dissertação no fim de 2024, mas só conseguiu analisar dados até 2017 porque não tinha dinheiro para fazer a viagem e acessar os dados mais recentes no Sedap. “Era um estudo bem legal de trajetória, ela fez até 2017 e disse: ‘Olha, queria continuar, mas não vou porque não tenho condições financeiras’”, lembra.

Além desses impactos apontados, as próximas sessões deste estudo da Open Knowledge Brasil se dedicam especificamente a abordar tecnicamente as mudanças implementadas nas bases de dados, evidenciando as questões de pesquisa que deixaram de ser respondíveis com o novo formato de divulgação de dados educacionais adotado pelo Inep.

Referências

1. Disponível em: <<https://scholar.google.com.br/?hl=pt>>.
2. Disponível em: <<https://scholar.google.com/intl/en/scholar/about.html>>.
3. Disponível em: <<https://www.abraji.org.br/cursos/4-curso-on-line-de-jornalismo-de-dados-com-sql>>.
4. Disponível em: <<https://escoladedados.org/courses/analise-de-dados-educacionais/>>.
5. Disponível em: <<https://escoladedados.org/courses/jornalismo-de-dados-para-coberturas-locais/>>.
6. Disponível em: <<https://www.youtube.com/watch?v=S4MRq1ligSA>>.

PRODUTO

1

